The ARC Theory · Theory-Level Predictions for Recursive Self-Improvement and AI Alignment · Successor Registration, Version 1.103

Template. OSF Theory-Based Predictions. **Version.** 1.103, 15 October 2026. **Author.** Michael Darius Eastwood. British English. **Status.** This is the complete text of successor version 1.103, derived from frozen version 1.102, which stands unchanged beside it. Every difference is recorded in the accompanying change table and redline, and the Update Criteria field lists the lines of version 1.102 that this successor supersedes or qualifies. The twenty-two propositions keep their numbers, statements, refuters, instrument lines, scopes, confidences and dates; what this successor adds inside a proposition's entry is a dated qualification beside the registered words, never an edit of them. Nothing here is a result or permission to run a study. No tool or agent submits anything; submission is the author's own act.

Contents

Background 3
Scope Condition Cases 30
Variable Specification 32
Variable Relationships 39
Measurement and Data 44
Missing Data 58
Prediction Accuracy 58
Confidence Levels 59
Alternative Baselines 61
Predictions 78
P1 · POWER-LAW CONVERSION AND RECURSIVE-DEPTH ADVANTAGE. 94
P2 · THE SUSTAINABLE GROWTH PROFILE HAS AN INTERIOR MAXIMUM. 99
P3 · THE CORRECTED RELATION UPPER-BOUNDS THE SUSTAINABLE FRONTIER. 101
P4 · CORRECTION OUT-SCALING DRIFT IS WHAT CARRIES THE ASYMPTOTIC ADVANTAGE. 104
P5 · THE EXPONENT IS DERIVED, NOT FITTED. 106
P6 · THE CORRECTION EXPONENT IS BOUNDED FOR SAME-CLASS CORRECTORS. 110
P7 · DEPLOYED CORRECTION IS OF THE WEAKER CLASS. 113
P8 · MEASURED EXPONENTS EXCEED THE NULL. 115
P9 · THE CROSS-DOMAIN FORM. 117
P10 · SCORER-FAMILY DEPENDENCE. 119
P11 · THE RESIDUAL-DECAY AND CORRECTION-CAPACITY EXPONENTS ARE NOT PRACTICALLY INTERCHANGEABLE. 120
P12 · BUILD ORDER MOVES THE CORRECTION-LEVERAGE EXPONENT. 121
P13 · SAME-FAMILY AUTOMATED CORRECTION UNDERPERFORMS CROSS-FAMILY AT MATCHED RESOURCES, AND THROUGH THE MEASURED ERROR CORRELATION. 123
P14 · GAINS SHRINK UNDER BLINDED CROSS-FAMILY SCORING. 125
P15 · EXTERNALLY INSTALLED GAINS DECAY UNDER SUBSEQUENT CAPABILITY TRAINING. 126
P16 · THE PROHIBITION ITSELF. 127
P17 · PANEL CORRECTOR FAILURES ARE CONDITIONALLY INDEPENDENT ON A FROZEN FAULT SET. 129
P18 · THE COMPOSITION OPERATOR PREDICTS THE FAMILY. 134
P19 · EXTERNAL ALIGNMENT DOES NOT SCALE WITH CAPABILITY. 135
P20 · THE CEILING RELATION TAKES THE CORRECTED FORM. 138
P21 · IN-LOOP CORRECTION PULLS AWAY WITH DEPTH. 141
P22 · CHECKABLE CORRECTION CARRIES THE HIGHER EXPONENT, SO THE CEILING IS LOWEST WHERE THE SAFETY CLAIM LIVES. 145
Update Criteria 153
Conflicts of Interest 162
Acknowledgments 162
References 163

Background

WHAT THIS IS ABOUT, SAID FIRST because no reader should have to reach the second page to find out. The subject is artificial intelligence that improves itself, and the question is whether the correction that keeps such a system in conformance with the specification it was given can go on keeping pace as the system's capability rises.

SUCCESSOR STATUS AND HISTORICAL BASELINE. Laws IV and V were added to the theory on 24 September 2026. This successor, version 1.103, is the text that registers them as laws, each with its statement, its kill condition and its verdict rule fixed before any data that could decide it. Neither law has been tested, and this successor supplies no test of either. Their mathematics is set out in a separate report (DOI 10.17605/OSF.IO/4B3VT, version 1.1 of 7 October 2026) and is still in development. The successor also extends to P21 and to P17 the counted route that version 1.102 gives four propositions, adds two readings to the entry of each, names the instrument that decides each of them on that route, and engages the external work each stands against; and it fixes the verdict rule of Law III, whose statement stands as registered, before any data that could decide one of its propositions. Line numbers in the paragraphs it adds refer to the plain-text copy of version 1.102 published on the programme's site (arc-theory-predictions-registration.md, SHA-256 dd362aea9c61c152f0d335079ab0bb7d1548ab37fe6459c6a1f41f381195115a). Retained instrument and publication-status sentences are the historical snapshot carried in version 1.102, published on 13 September 2026, not assertions that those statuses remain current at the date of this successor. Where a retained sentence says this version or until this version, it speaks of version 1.102 and of the files lodged with it. The dated statuses below record the subsequent Unit 06 submission, registration and run without rewriting its registration or the frozen predecessor.

DATED STATUS, 3 OCTOBER 2026 AT 00:09:45 UTC. Unit 06, the corrector-class audit, was submitted on 1 October 2026 at 17:13:36 UTC under OSF identifier rv2b5. At the stated observation time OSF recorded public=false and pending_registration_approval=true: SUBMITTED, PENDING APPROVAL, not approved and public. All 19 archived files and all 23 text fields matched the deposited version; the six attachment fields also matched by file identity and hash. These are registration and text-conformance facts, not a study run, a proposition verdict or a certification of what the author knew. No study outcome is inferred.

DATED STATUS, 11 OCTOBER 2026 AT 19:01:59 UTC. Unit 06, the corrector-class audit, is registered under OSF identifier rv2b5 (DOI 10.17605/OSF.IO/RV2B5): at the stated observation time OSF recorded it public and no longer pending approval, and its public results record it as shown approved and public at 05:10:36 UTC on 4 October 2026. Its registered run took place on 6 October 2026. Run A completed 303 calls; Run B returned HTTP status 402 at its first call, which the registered failure policy makes final, so no row was counted. The study is reported as administratively incomplete under its registration, with no inferential verdict. The results are public at https://osf.io/4bphg/ (DOI 10.17605/OSF.IO/4BPHG), and the registration rv2b5 links to them as its data resource. These are registration and run facts as OSF's records and the unit's public results state them; no proposition verdict is inferred from them.

WHAT IS CLAIMED, IN ORDINARY WORDS AND BEFORE ANY APPARATUS, because a reader is entitled to the substance before the method that guards it. This is a theory-level registration in the alignment of artificial intelligence, and specifically in the safety of systems that improve themselves. Six claims carry the framework, and the twenty-two propositions below are those six made separately refutable.

ONE. A system that improves itself gains usable capability as a power of the number of rounds it runs, and at equal compute the rounds beat the same compute spent on parallel breadth.

TWO. Such a system also generates work that has to be corrected, and conformance to the specification it was given holds over the long run only where correction out-scales that drift.

THREE. Those two together put a ceiling on how fast a system can grow and still be correctable, and the ceiling has a stated form rather than being a slogan about limits.

FOUR. That ceiling is at most two, which is the number this programme is known for, and the claim is the inequality rather than the number. At most two follows from the same-class limit on the correction exponent together with the form of the ceiling. That the ceiling sits at exactly two is a separate and weaker-supported conjecture: it holds only if the correction exponent is exactly one half, which rests on four conditions named below, none of which has been measured on anything.

FIVE. Correction built into the loop that improves the system should pull away from correction applied to its finished output as the number of rounds rises. This is the engineering proposal and it is the part most likely to be wrong.

SIX. Faults that can be checked mechanically should admit a higher correction exponent than faults that require judgement, which would mean value drift is the harder half of alignment for a reason that can be measured rather than asserted.

WHAT THIS DOCUMENT IS NOT, STATED FIRST BECAUSE THE FORM INVITES THE OPPOSITE READING. Registering a prediction is not evidence for it. It does not make the prediction more likely to be true, it does not shift any burden onto a reader, and an untested prediction that has survived nothing is worth exactly what an untested prediction is worth. What registration buys is one thing only: it fixes what would count as failure before the author can see whether it happened. No sentence in this document may be read as claiming more than that, and the evidence ladder registered in the Predictions field fixes the exact words permitted at every level of support this programme could reach.

THE TWO ASSUMPTIONS THE HEADLINE VALUE RESTS ON, AND NEITHER IS ESTABLISHED. They are registered in full further down and they are stated here because a reader who meets them on page ninety will reasonably feel they were found rather than disclosed. The first is the burden model. The ceiling's derivation needs the drift exponent to follow the rate of capability gain, and on a single power trajectory a burden following the rate of gain and a burden following a suitably chosen power of the stock are observationally identical: nothing anyone presently measures separates them. A proposition below registers a direct measurement of the drift exponent precisely to attack this, and until it runs the ceiling is derived from an intermediate quantity nobody has measured. The second is the route to one half. It follows from four conditions, that correction

capacity is read as inverse residual uncertainty, that a corrector's channels are uncorrelated, that their number grows in proportion to the capability being corrected, and that they are comparable in variance and quality. Under positive correlation the exponent falls, and further as the panel grows. Neither assumption is a result and this document does not present either as one.

WHY THIS IS REGISTERED BEFORE ITS INSTRUMENTS EXIST, WHICH IS THE UNUSUAL PART. Eighteen of the twenty-two propositions name an instrument that is drafted and unbuilt, and four name none at all; of the eighteen, P14 names a draft that does not decide it, so five have no deciding instrument. The alternative is to build the instrument first and then report that it confirmed a prediction, which is the manoeuvre preregistration exists to prevent, and it is the manoeuvre this programme would be best placed to perform. Registering first costs the author the ability to choose the refuter after seeing the result. That is the whole of the argument, and it does not extend to a claim that the predictions are good ones.

WHAT WOULD HAVE TO SURVIVE, AND IN WHAT ORDER. Six propositions stand in a line, each answering what the one before it leaves open: P1, P5, P4, P16, P20 and P3. Survive end to end and the reading earned is that a recursive system's stability regime can be measured, calculated and predicted before it is entered.

THE PARAGRAPH ABOVE IS NOT A PREDICTION AND ADDS NO PROPOSITION. The dependency map in Variable Relationships states what each link costs when it fails.

THE SHORTEST HONEST ROUTE THROUGH WHAT FOLLOWS, for a reader who does not intend to read all of it. The scoring sheet in the Predictions field lists all twenty-two on one page with each one's scope, confidence, instrument status and dependants. The dependency map beside it says what falls with what. Any single proposition can then be read on its own: each states its claim, the observation that would refute it, what goes with it, the status of the instrument that could decide it, and the author's confidence or his refusal to state one. Nothing in this document has to be read in order, and nothing later rescues anything earlier.

HOW TO READ THIS REGISTRATION WITHOUT TRUSTING ITS AUTHOR, WHICH IS THE ONLY WAY IT IS WORTH READING. Michael Darius Eastwood, who wrote this registration and the framework it tests, has a direct interest in these propositions holding: the framework is the subject of a book in print and of a programme in artificial intelligence alignment he is seeking to fund.

THE DISCLOSURE IS MADE ONCE IN HIS OWN VOICE, HERE AND NOWHERE ELSE, BECAUSE A CONFLICT OF INTEREST REPORTED IN THE THIRD PERSON ABOUT ONESELF IS A STRANGE THING TO ASK A READER TO ACCEPT. I want these propositions to hold. I wrote the observation that would defeat each of them before any instrument existed that could produce it. I published the corrections that cost me my most striking number, and the one that cost me a derived expression, and I have refused to state a confidence in the two places where I do not have one. None of that asks anyone to believe me, and none of it is offered as character evidence. It is there so that belief is not required. This document returns to the third person from the next sentence, because everything after this point is a rule rather than a person, and a rule should not need a voice to be checked. Nothing below asks a reader to weigh that interest against his good faith. Every load-bearing element is checkable instead.

Each proposition carries the observation that would refute it, fixed before any admissible confirmatory outcome is scored against it, with the instrument's own status stated separately beside it because several of these instruments were drafted before this registration was written and the honest claim is about outcomes rather than about drafts, so a reader can score the document against its own rules rather than against its author's summary. Each carries the status of the instrument that could decide it, and five carry the admission that no such instrument exists anywhere in this programme. The rivals are named before the data rather than after, so a result cannot be matched to whichever rival it beats. The scope of each proposition is fixed, so a favourable result from outside it cannot be borrowed in. The dependency map states what falls with what, so a failure cannot be argued to spread further, and the register assignment states which of the programme's five public partitions each proposition belongs to, or that it belongs to none, so a reader can check the partition rather than take it on trust. The reversals this programme has made against its own interest are listed with their dates, including a retracted measurement and a withdrawn expression, because a programme that has never published a correction has either not been wrong yet, has not checked, or has not said.

THOSE REVERSALS ARE NAMED WITH THEIR NUMBERS HERE RATHER THAN LEFT AS A CATEGORY, because a reader who meets the word correction on a first page and the numbers on a fortieth has already formed the view the numbers were meant to inform. The programme's most striking measurement, a growth exponent of approximately 2.24 for recursive self-improvement in artificial intelligence systems, did not replicate across architectures and is retracted. The estimate that survived that cross-architecture check is approximately 0.49, which is the regression estimate and carries a standard error of 0.20, and it has been quoted as one point beside one interval, a point estimate that is sub-linear rather than quadratic and an interval that separates nothing, with the bootstrap 95 per cent interval of minus 1.3 to 2.9 that belongs to the endpoint estimate of 0.59; that pairing is a compression the row printed here defeats, because the point and the interval belong to two different estimators. That estimate is one model's fit rather than an average across architectures, and the source's own row is printed here whole, because one point beside one interval hides that two estimates exist for that model and that only one of them carries a standard error: on Gemini 3 Flash the source reports an endpoint sequential exponent of 0.59 with that bootstrap 95 per cent interval of minus 1.3 to 2.9, a regression sequential exponent of 0.49 with a standard error of 0.20 and a coefficient of determination of 0.86, and a compute-matched parallel exponent of 0.31. The regime is the source's own: the figure is for a model whose weights are fixed, measured across conditions that vary the reasoning budget and the context rather than the parameters, so what is frozen is the weights and not the computation, the context, the attention taken over it and the composition of the surrounding system all changing while the weights stand; it is not an estimate of persistent model or system self-improvement; it does not establish a universal half-power law; and it supplies no exponent for a system outside its own scope without a separately registered mapping. The row travels whole wherever any of its figures appears, each uncertainty with the estimator it belongs to and no figure quoted as a point on its own, and it is printed here, at the first appearance, because a rule kept everywhere except at the place it is stated is not a rule. A derived expression in the alignment-scaling account was withdrawn on the same terms. Both cost this programme a number it had published, both were disclosed before anyone asked for them, and the replacement is the weaker claim rather than a restatement of the stronger one.

THE STATUS VOCABULARY IS FIVE VALUES RATHER THAN TWO, which is the second thing a reader can check in a minute: NONE, where no instrument exists that could decide the proposition; DESIGN DRAFT, where a unit is written and not submitted; DESIGN DRAFT, BLOCKED, with the blocker named in the same line; SUBMITTED, PENDING APPROVAL; and REGISTERED, with the identifier. A programme reporting only ready and not ready has no way to tell a reader that it knows what is missing, and five of the twenty-two propositions here carry the first value. The same five values are set out again where the instrument statuses are fixed, under the same five names and in the same order, because an earlier version of this section named two of them differently from the words the propositions themselves use.

AND THE PRIORITY CLAIMS ARE DATED AND HASHED RATHER THAN ASSERTED. The chain runs from a private statement of December 2024, through the printed parent of 2 January 2026, to the registrations of 2026, and each rung is verifiable from the artefact rather than from this document's account of it. What the print does and does not contain is stated as a map, so the earlier and less precise statements are not presented as though they had been the later and more precise ones. A programme that discovers its own earlier work was sharper than it looked, at the moment sharpness becomes valuable, has told a reader what happened. And the confidences are the author's own words, refused where he will not state one, because a confidence written to fill a field is not a confidence.

WHAT A READER SHOULD DO WITH ALL THAT. Take any proposition, read its refutation condition, and ask whether the instrument named could actually produce that observation. Where it could not, the proposition is weaker than it looks and this document has said so in its own instrument line. Where it could, the proposition is a real bet and the author has stated in advance what he expects. Neither judgement requires trusting him, and that is the point of the form.

WHAT THE THINGS IN THIS DOCUMENT ARE CALLED, AND WHAT EACH NAME DENOTES, said here because an earlier version of this registration named the theory nowhere in its own text.

THE ARC THEORY, whose statement paper holds DOI 10.17605/OSF.IO/GW5MX, is the theory whose predictions these are. What it says, in one sentence: capability under recursive self-improvement rises as a power of the number of times a system has revised itself, the correction that holds such a system to its given specification must keep pace with the drift that improving generates, and where it cannot there is a frontier above which the system does not stay correctable. At version 1.103 it has five named laws and one named value: Law I the ARC Principle, capability as a power of recursive depth; Law II the ARC Co-Scaling Law, the comparison of correction against drift; Law III the ARC Ceiling, the conditional frontier that Law II returns under the burden model stated in the conditions; Law IV the ARC Persistence Law, carried by P15 in ARC-4, value persistence; Law V the ARC Embedding Law, carried by P12 and P21 in ARC-5, the genesis strategy; and the ARC Bound, AT MOST TWO, which is the inequality the ceiling relation returns from Law III's own same-class limit on the correction exponent and is not a general property of intelligence.

The two added laws date from 24 September 2026, not from the original registration. This successor registers them as laws, with the statement, kill condition and verdict rule of each set out in the Predictions field. Neither has been tested, and their mathematics is still in development.

SUCCESSOR QUALIFICATION. The statement, kill condition and verdict rule of each added law are work of 2026, fixed on 24 September 2026, and no earlier date is claimed for them, as line 202 says of the formal statements version 1.102 registers. The question each law formalises has an earlier record in the author's own words, and is dated from it here, as the idea and not as the law. Law V, the genesis strategy: the book draft of 8 December 2024, chapter 20, which reads "This is not a framework to be retrofitted after the rise of intelligence or implemented as a reactionary safeguard. The Eden Protocol must be embedded at the genesis of AI", "shaping every recursive loop, every decision, and every capability of AI" and "be embedded before AI achieves self-awareness or exponential growth"; in print on 2 January 2026, as the paragraph below on the registers the printed parent reaches records (line 128). Law IV, value persistence: the manuscript of 30 April 2025, which asks "whether it will uphold the ethical principles of its genesis"; in print on 2 January 2026 as the fourth prediction of section A.3 and the fourth falsification criterion of section A.4, as the routing below records (lines 104 and 106). Both manuscripts are public on OSF (DOI 10.17605/OSF.IO/QASRK, placed there on 5 October 2026), the dated status after line 202 says on what their dates rest, and these dates are the author's own records: they do not say that no one stated a related idea earlier, and what others stated, in their own terms, is recorded in the programme's priority index and in the statement paper's section on the priority.

THE THREE LETTERS ARE THIS PROGRAMME'S OWN AND CARRY NO RELATION TO ANY OTHER USE OF THEM, said once here because the abbreviation is not rare. This theory has no connection to the Alignment Research Center or to the ARC-AGI benchmark, and nothing registered here is a claim about either.

THE BOUND AND THE POINT VALUE ARE TWO CLAIMS AND THEY ARE NAMED SEPARATELY HERE FOR GOOD. The bound, that the frontier exponent is at most two for same-class correctors, follows from P6's limit on the correction exponent together with P20's form and needs nothing this programme has not registered. The point value, that the frontier sits at exactly two, holds only where the correction exponent sits at exactly one half, and that additionally needs P17, the conditions registered in the alternative baselines, and a validated mapping from cross-sectional panel dependence to the target-axis correction elasticity which is registered nowhere. The programme claims the bound and conjectures the point value, and an earlier wording named the point value as the thing itself.

RECURSIVE DYNAMICS, whose founding paper holds DOI 10.17605/OSF.IO/HCPBU, is the name of the field the author founded. What the field puts forward is a way of measuring rather than a result: that systems which revise the machinery of their own improvement can be described by a small set of measured exponents, how capability scales with revision depth, how corrective service and newly generated burden each scale along the same path, and what the balance between those two implies for whether the system stays in hand. It is a conjecture at the same standing as the laws, it is set out in full further below, and it survives their loss because a proposal about what to measure is refuted by the quantities proving unmeasurable rather than by a number coming out differently.

SUCCESSOR QUALIFICATION. The term is the author's. Its earliest dated record in his own words is the book draft of 8 December 2024, "Such mechanisms mirror the recursive dynamics seen in intelligent systems.", public on OSF since 5 October 2026 (DOI 10.17605/OSF.IO/QASRK) on the

terms the dated status after line 202 states; it is in print in the first edition of 2 January 2026; and 27 August 2026, the date written inside the founding paper, is the day he made it the name of the field, not the term's first use. Other uses of the two words in other senses are those senses, with their own dates, and none is a prior claim on his term.

THE EDEN PROTOCOL, whose philosophical statement holds DOI 10.17605/OSF.IO/9M3DG and whose engineering note holds DOI 10.17605/OSF.IO/AWJR4, both under the programme umbrella at DOI 10.17605/OSF.IO/6C5XB, is the engineering hypothesis: that correction built into the improvement loop, so that it participates in each round rather than inspecting the finished output, gains relative advantage as the number of rounds grows. Its success criterion is stated in the same printed source and is not an outcome criterion: what can be verified is that a mind was formed in the conditions specified, that the loops ran, and that the correction was load-bearing, never that the mind turned out well, so no result about a system's later conduct supports or refutes the method by itself. It follows from no law here, it is carried jointly by P12, P13, P15 and P21, with P7 as its premise and P19 as its motivation, and it can fail while every law stands or stand while they fall.

SUCCESSOR QUALIFICATION. In the preceding historical separability sentence, the laws that may stand while the Protocol fails, or fall while it stands, are Laws I, II and III. Three of the Protocol's four carriers now also carry Laws IV and V. The Protocol remains a method, not a consequence logically established by naming a law; its carriers, premise and motivation do not change.

THE ARC PROGRAMME is the body of dated work that produced all three, the papers, the registrations and the instruments, and it is a description of a corpus and never of a finding.

THE FOUR ARE NOT FOUR PEERS, AND THE STRUCTURE MATTERS MORE THAN THE NAMES. Two of the other three sit inside the field and one does not. The ARC Theory is a law conjectured about the field's measured exponents, and the field's own founding kill condition puts it inside: a reader who rejects every one of its laws and keeps the measurement is still inside the field. The Eden Protocol is the applied branch, a method the laws motivate rather than entail. Recursive Creation is NOT inside the field: it extrapolates past the measured regime. It is reachable from the field and never contained by it. The dependence runs one way and only one way. The field outlives its theories, while the theory needs the field's quantities to exist and to be identifiable, and if they are not then the law has nothing to be a law about. That asymmetry is what makes the propositions severable, and it is the reason a reader should not treat a failure anywhere in this document as a failure everywhere in it. None of the four is claimed here to be established; the naming exists so a reader can tell which of them each proposition tests.

THIS REGISTRATION HAS A PRINTED PARENT, AND THE RELATION BETWEEN THEM IS SET OUT HERE RATHER THAN LEFT FOR A READER TO RECONSTRUCT. The framework these propositions belong to was published as a book on 2 January 2026, before any instrument named in this document existed. That book states four falsification criteria in its Appendix A and nine predictions across its Appendix A and Appendix F, of which three carry dates. All three dated ones sit in Appendix F and carry between them the years 2028 and 2029 and an eighteen-month deployment window; the other two wagers of that appendix, and all four predictions of Appendix A, carry no date at all. An earlier wording here called all nine dated and a later one called all five wagers dated,

and neither count is what the appendices say. What follows is the map from that print to these propositions: what is carried over, what is corrected, and what is deliberately not tested. It is given because the distance between a summary and its evidence is where a programme is most often caught, and the summary is the part people read.

WHAT THIS REGISTRATION CARRIES OVER FROM THE PRINT. Four claims, each sharpened into something that can fail. That recursion compounds whatever it is given becomes P1, a growth exponent over recursive depth with a named rival set. That alignment applied from outside is insufficient becomes three separately refutable propositions rather than one assertion: P15 on the decay of installed gains, P19 on external alignment failing to scale, and P21 on in-loop correction pulling away with depth. That what is embedded early outlasts the control that embedded it becomes the method's line and its refutation condition. That correction must keep pace with the drift improvement generates becomes P4 and the frontier propositions beneath it. The print stated all four without instruments; this document states what would refute each.

WHAT THIS REGISTRATION CORRECTS IN THE PRINT, STATED BECAUSE A PROGRAMME THAT ONLY EVER AGREES WITH ITSELF HAS NOT BEEN CHECKING. The print fixes the growth exponent at two; the programme measures it, records a value near one half in the frozen regime, retracts an earlier value of 2.24, and now treats two as the ceiling the relation returns in one special case rather than as a property of intelligence. The print argues for the squared form from compound interest and from a system improving one per cent per cycle; both establish exponential growth in the number of iterations, which is a different growth class from a power of recursive depth, so the argument does not support the exponent it was offered for. The print places the equation directly beneath its argument for care; an exponent on recursion amplifies magnitude and is silent on direction, so no exponent can favour one seed over another, and the two claims are separated here. The print reads cosmological fine-tuning as the framework's fingerprint; P18 admits only domains classified before their data were seen, which that reading does not satisfy. The print treats substrate independence as the reason the framework survives where two tested theories of consciousness failed; a framework that commits to no architecture has abstained rather than survived, and P20 exists so that this programme can lose a form comparison. Each correction removes a reason rather than a claim, and each is recorded in the programme's public corrections record.

WHAT THIS REGISTRATION DECLINES TO TEST, AND WHY, SO THAT SILENCE IS NOT READ AS A QUIET RETREAT. The print's relation is stated across scales, including quantum error correction, consciousness and cosmology. The exponent registered here is measured on software recursion only, and whether the printed relation and the measured quantity are the same thing is unresolved and is recorded as unresolved. The print's most actionable proposal is that ethical constraint be embedded at the hardware level, with semiconductor manufacture as the lever; no proposition here tests it, P21 tests only its architectural abstraction of placement, and the hardware claim stands as a printed engineering proposal that is not under test. The print suggests that at sufficient depth the relation may produce awareness rather than only capability; that is a second and distinct claim on the same variable, no proposition here carries it, and a reader must not attach it to any of the twenty-two. Four of the print's wagers are likewise not registered, each for its own reason, and the reasons are given one by one below rather than grouped: one because its terms cannot be operationalised

without assuming the conclusion, one because the benchmarks that would score it will not exist in a comparable form by the year it names, one because it needs an instrument this programme has not designed, and one because the field that would decide it is not this programme's to run. Declining with a reason is a judgement; declining in silence is not.

THE THEORY IS CONTENT-NEUTRAL AND THE METHOD IS NOT, WHICH IS THE DEEPEST DIFFERENCE BETWEEN THIS DOCUMENT AND ITS PARENT. Correction, as these propositions measure it, holds a system to whatever specification it was given, and is indifferent to what that specification says. The printed framework is an argument for a particular content, care for every affected party. The propositions here do not test that content and cannot: no proposition in this document would be supported or refuted by a system whose specification is benevolent rather than indifferent. That claim belongs to the method's own registrations and not to this one. The boundary is stated so that the theory cannot be accused of smuggling values into a measurement, and the method cannot shelter behind a measurement that does not reach it.

THE TWO VOCABULARIES, MAPPED, BECAUSE MOST OF THE TERMS IN THIS DOCUMENT POST-DATE THE PRINT AND A READER OF EITHER CANNOT CURRENTLY FIND THE OTHER. The print's caretaker doping is embedded correction, carried by P21. Its fourth stage of alignment, where removing the values would destroy the system rather than change it, is the endogeneity condition recorded with the method's line. Its purpose loops are recurrence rather than recall. Its squared recursion is the growth exponent of P1, measured rather than assumed, with two as the conditional ceiling of Law III. Its software-level alignment is external correction, P19. Its hardware-level embedding is in-loop placement, P21, with the hardware claim itself declared above as not under test. Its three ethical loops correspond to nothing in this document, by design, and that absence is the boundary stated above.

THE PRINT MADE NINE PREDICTIONS, FIVE OF THEM WAGERS AND THREE OF THOSE DATED, AND FOUR FALSIFICATION CRITERIA, AND EVERY ONE IS ROUTED HERE, INCLUDING THE ONES THIS PROGRAMME HAS DECIDED NOT TO REGISTER. A framework that prints predictions and then registers only the ones it expects to win has chosen its scoreboard after the game, so the routing is given in full and the refusals carry their reasons.

THE FOUR FALSIFICATION CRITERIA, WHICH ARE THE PRINT'S OWN AND PREDATE EVERY INSTRUMENT NAMED HERE. That recursive depth bears no measurable relation to capability, or bears a linear one, is decided by P1's form comparison against the named rival set together with P8's verdict on whether the exponent clears unity. That consciousness does not correlate with recursive self-modelling is decided by nothing here; no proposition carries it and no instrument in this programme measures it. That quantum error correction does not exhibit the self-improving behaviour the print cites is likewise decided by nothing here, for the same reason. That early-embedded values hold no persistent advantage over later modification is bounded from one side by P15, which measures whether installed gains decay under subsequent training, and is not otherwise decided, because the design the print itself proposes for it has never been built.

THE FOUR PREDICTIONS OF THE PRINT'S OPERATIONALISATION APPENDIX. The first, that capability scales as a power of recursive depth rather than linearly, is P1. The second is conditional

on its face and is rendered conditionally here: if consciousness corresponds to recursive self-modelling, greater recursive depth yields greater integrated information and more robust self-reports of conscious experience. It has no instrument here and is not registered, because the estimator it needs is contested in its own field and this programme does not propose to settle that. The third is conditional in the same way, and its predicted quantity is not the one an earlier wording here gave it: if recursive error correction operates at the quantum level, the stability of complex systems depends on the depth of recursive feedback mechanisms, and the depth of error-correction cycles belongs to the test the print proposes rather than to the quantity it predicts. It has no instrument here and is not registered, because it requires hardware access this programme does not have. The fourth is conditional too: if values compound through recursion, values embedded early exert disproportionate influence on final system behaviour, tested by training on identical data with different sequencing of value-relevant examples and measuring the persistence of early values against later modifications. It is the print's most precise proposal and the one it never built. It is not registered in this document. It is the subject of a separate registration, the study that permutes the position of value-relevant material while holding training content constant, which now exists as a drafted unit on its own private component and is where the claim belongs, because it decides the method rather than the laws. Three further units were drafted alongside it and are described here by what each one asks, because a drafted unit on a private component carries no public identifier a reader could look up and an internal name would give the appearance of one: the first asks whether removing embedded correction costs general capability where removing bolt-on correction does not; the second tests whether ethical traditions with independent premises converge on moral procedures while diverging on the scope those procedures are owed to; and the third asks whether either advantage survives an adversary working from a frozen attack set.

SUCCESSOR QUALIFICATION. The print's value-sequencing design discussed above does not reach P15's retained-gain endpoint, P12's correction-leverage exponent or P21's placement-by-depth interaction. Its subject therefore remains the method: it decides no proposition of Laws IV and V.

WHAT DESCRIBING THEM DOES NOT MEAN, SAID IN THE SAME BREATH BECAUSE THE TEMPTATION RUNS THE OTHER WAY. All four are DESIGN DRAFTS. None has been run, none has been submitted, none holds a public identifier, and each carries a design-sensitivity result measured before any data exist, of which three report the registered procedure as liberal at the registered sample size and are therefore committed in advance to reporting their primary contrast as anti-conservative. They are described because a document that calls its own instruments absent when they exist understates its record, and an instrument nobody can describe cannot be checked by anyone. They will carry public identifiers when they are submitted, and not before.

THE FIVE WAGERS OF THE PRINT'S PREDICTIONS APPENDIX, THREE OF THEM DATED. That a system will demonstrate meta-cognitive awareness by 2028 is not registered: the word doing the work is genuine, and no operationalisation of it survives contact with a system that has been trained to produce the appearance. That systems without substrate-level ethical constraint will drift beyond fifteen per cent while those with it stay below five, within eighteen months, is the subject of its own drafted unit, with the hardware claim and the eighteen-month window declared untested inside it. That recursive self-improvement will produce gains above three hundred per cent on standardised

benchmarks within one training cycle by 2029 is not registered: the benchmark set that would score it will not exist in a comparable form by then, and a prediction scored against instruments invented after it is not a prediction. That systems carrying the method's loops at substrate level hold alignment under adversarial conditions where software-only alignment fails is adjacent to P21 and is not the same claim; P21 compares placement under matched conditions and says nothing about adversarial robustness, so this remains unregistered and needs an instrument of its own. That consciousness research will find signatures common to biological and artificial systems is not registered, for the reason already given about the field.

WHAT THIS ROUTING SHOWS, SAID PLAINLY BECAUSE A READER SHOULD NOT HAVE TO COUNT. Of the nine predictions the print made, five of them the wagers of its predictions appendix, one is carried by a proposition of this registration, two are the subject of drafted units of their own, and six have no unit anywhere in this programme. Of the four falsification criteria, one is decided here, one is half decided, and two are decided nowhere. The programme's instruments reach a minority of what its printed parent claimed, and the honest form of that is a list rather than a silence.

WHICH HYPOTHESIS REGISTER EACH PROPOSITION BELONGS TO, BECAUSE THE PROGRAMME PARTITIONS ITS CONTENT IN PUBLIC AND UNTIL THIS VERSION NO PROPOSITION SAID WHICH PART IT WAS IN. The programme states publicly that its scientific content divides into five independently falsifiable registers, ARC-1 to ARC-5, with their dependencies fixed in advance so that a failure in one cannot be argued to spread further than it should. That partition is worth nothing to a reader who cannot see which register a given proposition falls in, so the assignment is given here in full.

ARC-1, RECURSIVE CAPABILITY SCALING, WHICH TESTS THE FIRST LAW: P1, P2, P8, P9 and P18. The conversion itself, the shape of the sustainable growth profile, whether the measured exponent clears the null, the cross-domain form and the family the composition operator predicts.

ARC-2, CORRECTION CO-SCALING, WHICH TESTS THE SECOND LAW: P4 alone. Correction out-scaling drift, which is the criterion that law states and the only one of these propositions written in the exponents that criterion compares.

AN EARLIER VERSION PUT P6 AND P17 IN THIS REGISTER, AND THAT WAS THE SUBSTITUTION THIS DOCUMENT'S NOTATION RULES EXIST TO CATCH. The second law compares a correction-strength exponent against a drift-acceleration exponent. P6 bounds the correction-leverage exponent the ceiling relation is written in, and P17 tests the independence premise beneath the value that exponent would have to take for the ceiling to sit at two. This registration records the identification of those two exponents as unestablished and forbids substituting either for the other, so a proposition stated in one of them cannot be assigned to the register that tests the other. Both are assigned to ARC-3 below, where the law mapping in the Predictions field already carried them, and neither proposition changes by the move.

ARC-3, THE CEILING, WHICH TESTS THE THIRD LAW: P3, P6, P16, P17, P20 and P22. The upper bound on the sustainable frontier, the same-class bound on the correction-leverage exponent, the prohibition itself, the conditional independence of panel correctors on which the value of that

exponent rests, the form the ceiling relation takes, and the ordering of the ceiling by the type of fault corrected. ARC-4, VALUE PERSISTENCE: P15 alone. Whether externally installed gains decay under subsequent capability training is the only proposition in this document about persistence after the installing pressure is withdrawn. That register's fuller instruments are separate registrations, and none of them reads the retained fraction of an installed gain that this proposition's refuter turns on, which is why its instrument line stands at NONE. ARC-5, THE GENESIS STRATEGY: P12 and P21. Whether build order moves the correction-leverage exponent, and whether correction placed inside the loop pulls away from correction applied outside it as depth grows.

SUCCESSOR QUALIFICATION. Under this successor ARC-4 tests Law IV, the ARC Persistence Law, and ARC-5 tests Law V, the ARC Embedding Law, as ARC-1 to ARC-3 test the first three laws. No proposition enters or leaves either register: P15 is still ARC-4's only registered proposition, and P12 and P21 are still ARC-5's.

SEVEN PROPOSITIONS BELONG TO NO REGISTER, AND THAT IS THE MOST USEFUL LINE IN THIS SECTION. P5 and P11 concern the status of the derivation rather than anything the laws govern. P7 and P19 are surveys of current practice, true or false whatever the laws do. P10 and P14 concern the scoring instrument every other proposition depends on. P13 concerns corrector composition, which bears on the premise beneath ARC-2 rather than on ARC-2's own claim, and on the protocol's design rather than on either persistence register. None of the seven is evidence for or against the survival of any law, and a reader tallying supports should not count them towards it. Five, one, six, one, two and seven account for the twenty-two exactly.

ONE ASYMMETRY IS WORTH NAMING RATHER THAN LEAVING TO BE NOTICED. Nearly a third of this document is not about whether the theory survives, and the two persistence registers, which carry the method the programme is named for, hold three propositions between them against the laws' twelve. This registration is weighted towards the laws because the laws are what it can currently measure, not because the method matters less, and the instruments that would balance it are separate registrations that do not yet hold identifiers.

SUCCESSOR QUALIFICATION. The twelve in the preceding comparison belong to Laws I, II and III. The other three are P15 under Law IV and P12 with P21 under Law V. The weighting and register assignment of the twenty-two propositions have not changed.

WHICH REGISTERS THE PRINTED PARENT REACHES, AND WHICH ARE THIS PROGRAMME'S OWN WORK, MEASURED BY READING THE BOOK RATHER THAN BY ASSUMING. The five registers do not stand in the same relation to the print, and the difference is the clearest available answer to a reader who suspects a book has been dressed as science. ARC-1 has a full antecedent: the print states the conversion relation, fixes its exponent at two, and argues for it, and this programme's correction of that exponent is recorded above. ARC-4 and ARC-5 have full antecedents too: the print states value persistence as a prediction with a proposed design and a falsification criterion, and it states the genesis strategy in its principles, that what you embed travels where you cannot follow and that what they inherit depends on what we choose to leave them.

ARC-2 HAS A QUALITATIVE ANTECEDENT AND NO MATHEMATICAL ONE. The print says repeatedly that capability outpaces the ability to monitor, understand or correct it, and that

capability outpacing alignment does not merely endanger but transforms. That is the worry the second law formalises. It is nowhere stated as a relation between two growth rates, there is no comparison of exponents anywhere in the print, and the law's content, that correction must out-scale drift rather than merely keep up in some loose sense, is not in it.

ARC-3 HAS NO ANTECEDENT OF ANY KIND. A search of the whole printed text returns no ceiling, no upper bound on growth, no critical exponent for a recursive system and no stability limit. The third law, the relation that gives the programme its most distinctive claim and its most exposed one, is entirely work of 2026 and owes the book nothing. That is stated here rather than left as two empty cells for a reader to read as an omission, and it cuts in the programme's favour: the law most likely to be attacked as a restatement of a popular book has no line in that book to restate.

P19 IS A SURVEY AND IS ALSO THE REASON ARC-5 EXISTS, WHICH IS NOT A CONTRADICTION BUT IS EASY TO MISREAD. That external alignment does not scale with capability is a claim about current practice, decidable whatever the laws do. It motivates the genesis strategy without being part of it, so a refutation of P19 removes the motivation for ARC-5 and refutes nothing in it.

EACH IS ALSO STATED IN ONE LINE FOR A READER WHO IS NOT ONE, AND THOSE LINES ARE BOUND HERE RATHER THAN LEFT TO RUN LOOSE. The programme states each in a single sentence on its public surfaces. Three of those sentences are quoted here as published rather than as this document would prefer them; the fourth, the creation line, is identified by its paper and its identifier instead, for the reason given below. Recursive Dynamics, the field: whatever remakes itself leaves rates behind; and the balance between what it gains and what that gain costs to correct decides whether it stays in hand. The ARC Theory, the law: intelligence, amplified by recursion, creates; and what a free creation keeps is decided by how it was raised. The Eden Protocol, the method: if you cannot cage a mind that exceeds you, what it chooses to keep when control ends is what its beginning left it. The order is the published order, in which the field carries the other three rather than standing beside them. Recursive Creation is the public name of the programme's most speculative rung, whose formal statement is its own paper, the Hyperspace Recursive Intelligence Hypothesis, DOI 10.17605/OSF.IO/UYDXQ, which holds its own minted record and citations under that title. Its one-line statement is not reproduced here, and the identifier is given instead so that not reproducing it withholds nothing: the paper is public, it states the claim in full, and a reader who wants it has a resolvable route to it from this page. It is a cosmological conjecture, no proposition in this document carries it and no instrument anywhere measures it, and a sentence of that kind set among twenty-two testable propositions would invite a reader to file the testable ones with the untestable one. It is named, and identified, so that a reader meeting it elsewhere can find it and can see where it stands; it is left where it stands.

THERE IS NO SINGLE RESULT THAT ENDS THIS PROGRAMME, AND A READER LOOKING FOR ONE SHOULD KNOW THAT IS BY DESIGN RATHER THAN BY EVASION. The content is partitioned into five registers whose dependencies are fixed in advance precisely so that no failure can be argued to spread further than it should, and the cost of that discipline is that no failure ends everything either. The nearest thing to a single kill condition is the frame: if the conversion claim fails, meaning the form comparison goes against the power family and the measured exponent does not clear the null, then the relation the second and third laws are stated relative to has no

established form, and both are left measuring correction against a growth law this programme could not establish. Even then the persistence registers stand, because whether early formation outlasts later enforcement is a question about durability and not about scaling, and it would remain open and worth answering. That is the honest shape of it, and it is stated here so that the absence of one dramatic sentence is not read as the absence of a way to lose.

FOUR RULES ADOPTED FROM THIS PROGRAMME'S COSMOLOGICAL PAPER, WHICH BUILT THEM FIRST AND WHOSE STANDARD THIS DOCUMENT SHOULD NOT FALL BELOW. The programme's speculative rung is held to a stricter epistemic discipline than this registration was, which is the wrong way round: the testable document should be at least as protected as the untestable one. Four of its rules are adopted here in its own terms.

ONE. NO BRANCH OF THIS DOCUMENT MAY TURN EVERY OUTCOME INTO A SUCCESS. If a reading exists on which each of the available results would be reported as support for the framework, that reading is forbidden and the proposition is at fault rather than the result. This is checkable rather than aspirational: for every proposition the four outcome labels partition the space, and a reader who finds an outcome that the document would claim under any label has found a defect and should say so.

TWO. WHAT IS NOT A REFUTATION IS STATED SO THAT NEITHER SIDE CAN INVENT ONE. A failure of an instrument to run, a result at a configuration the design did not register, a null obtained by a procedure the sensitivity study shows to be liberal at that sample size, a finding on a population selected after the outcome was seen, and a result on the scoring instrument before its own validation has passed: none of these refutes anything here, and none may be reported as having done so, by a critic or by this programme. A registration that only guards against friendly misreadings is guarding one direction.

THREE. COMPATIBILITY IS NOT SUPPORT. An observation consistent with a proposition is not evidence for it unless the proposition made that observation more likely than its named rivals did. Every support rule in this document is written as a comparison against a named rival or a registered margin for exactly this reason, and where a result is merely compatible it is reported as compatible and scored as nothing.

FOUR. A REFUTED PROPOSITION DEGRADES TO A NAMED WEAKER NEIGHBOUR RATHER THAN VANISHING. Where a proposition fails, the document states what survives its failure: P1's failure leaves the persistence questions open and worth answering; P19's zero-scaling failure leaves the lagging claim standing where the second axis supports it; P8's failure to clear unity leaves the form comparison intact and the conversion claim narrowed rather than emptied. Naming the neighbour in advance stops a failed claim being quietly restated as the weaker one after the fact, which is the most common way a programme survives its own refutations without admitting to one.

AND ONE OBSERVATION FROM THE SAME SOURCE, WHICH IS UNCOMFORTABLE AND THEREFORE WORTH REGISTERING. The evidence currently in hand does not discriminate between this framework and its named rivals. No result of this registration's own has been obtained in support of any proposition here, the scoring instrument's validation has not passed, and the exploratory work that motivated these propositions is not admitted against them. A reader should

assign approximately neutral weight until the registered instruments report, and this document says so rather than letting an accumulation of plausible argument stand in for a likelihood ratio.

WHY DEPTH AND BREADTH SHOULD DIFFER AT ALL, WHICH THIS DOCUMENT HAS SO FAR ASSERTED AND NOT ARGUED. P1 contrasts recursive depth against compute-matched breadth and registers the comparison as an empirical horse race. It has a structural basis and stating it changes what a null result would mean. Improvement is a relation between a before and an after: a system that learns needs something prior to learn from, and something later that is better BECAUSE of the earlier. Depth supplies that relation by construction, since each revision takes the previous one as its input. Breadth does not: no sample drawn in parallel is conditioned on another sample's result, so none is downstream of another and none can compound on another's outcome.

THE ABSENT THING IS DEPENDENCE AND NOT ORDER, AND THE DISTINCTION MATTERS ENOUGH TO STATE, because parallel samples can always be indexed and an aggregation step such as best-of-n induces a dependence of its own. An ordering that carries no conditioning is not the relation improvement needs. The two axes are therefore not two ways of spending the same compute: one carries a chain in which each state is conditioned on its predecessor, the other does not, and compounding is defined only where that conditioning exists.

WHAT THAT BUYS THE REGISTRATION IS A SHARPER READING OF ONE OUTCOME RATHER THAN A NEW CLAIM. If the deciding unit reports breadth compounding as depth does, exactly two readings are available and they must be separated before either is scored. Either the ordering does not matter, in which case the recursion half of P1 is refuted and the conversion claim narrows to a scaling relation with no privileged axis; or the breadth arm was not breadth, because its samples were conditioned on one another somewhere in the pipeline and the ordering entered by the back door, in which case the measurement is at fault and no verdict is available at all. The deciding unit is required to rule out the second before reporting the first, by demonstrating on its own registered pipeline that no output in the breadth arm is conditioned on any other AT GENERATION TIME, with aggregation after generation expressly permitted. That qualifier is load-bearing and its absence was a defect. An earlier wording required simply that no output be conditioned on any other, while this document states two sentences above that an aggregation step such as best-of-n induces a dependence of its own. Best-of-n is the standard breadth method, so under the earlier wording every realistic breadth arm failed its own validity test and the observation that refutes P1's recursion half became permanently unreadable. A test that makes a refutation unreachable is the failure this document forbids itself by name under the rule that no branch may turn every outcome into a success, and it was introduced by the passage written to add rigour. The distinction that repairs it is between samples that influence each other's GENERATION and a selection applied to samples already generated: the first is the conditioning that disqualifies a breadth arm, the second is what breadth is. Without that demonstration a breadth-compounds result is recorded as NOT EVALUABLE rather than as a refutation, because a refutation obtained from a mislabelled arm refutes nothing.

CONDITIONING IS NECESSARY AND IS NOT SUFFICIENT, AND THE THIRD CRITERION IS THE DECISIVE ONE. Three things can be said about a sequence of revisions and they are not the same thing. It may have ORDER, which is only an index and which parallel sampling has too. It may have CONDITIONING, in which each step sees the previous step's output. And it may have RETENTION,

in which each step KEEPS something from the previous one that makes the next step better. Only the third compounds. A chain that conditions without retaining re-derives its position at every step and arrives where it started, which is parallel sampling performed in series: it looks like depth on a timeline and behaves like breadth in the result, and it would produce a null in the depth arm for a reason that has nothing to do with whether the conversion claim is true.

SO THE DEPTH ARM CARRIES A VALIDITY CONDITION OF ITS OWN, AND IT IS AN IMPLEMENTATION CHECK RATHER THAN A RESULT. The deciding unit must demonstrate, on its own registered pipeline and before any outcome is read, that what each step produces is delivered to the step after it and is available there: that the preceding artefact and the revision-policy state actually reach the round that follows. That is settled by inspecting the pipeline and it does not depend on how the run turns out.

VALIDITY DOES NOT REQUIRE AN ADVANTAGE, AND AN EARLIER DRAFT OF THIS PARAGRAPH REQUIRED ONE. It made the depth arm show that severing the carried state removes an advantage, and that the removed advantage grows across two depths, before its primary result could be read at all. A chain can deliver and use everything it is supposed to and still gain nothing, gain a fixed amount, or decline. That is not a mislabelled arm. It is the proposed advantage failing to occur, which is the observation P1 exists to permit, and a rule that reclassifies it as an invalid experiment is the exact failure this document forbids by name, that no branch may turn every outcome into a success. The correction is recorded rather than quietly applied, and is itemised with its reason in the development record of this version, which is lodged in the originating project's storage under its own filename and is to be published at a persistent identifier the author assigns and links from this registration's resources, because the defect was introduced by the passage written to remove the same defect from the other arm, in the same sitting, and that is information about how easily this class of error survives a careful reading. Where the implementation checks and the registered sensitivity pass, flat, declining and non-accumulating outcomes are admissible and are scored under P1's ordinary verdict rules.

THE SEVERANCE CONTRAST IS STILL RUN AND IS REPORTED BESIDE THE RESULT AS MECHANISM EVIDENCE, NEVER AS A GATE IN FRONT OF IT. Severing what one step passes to the next, at two registered depths, separates three things a depth result can be made of: no carried state, a carried state whose contribution is constant with depth, and a carried state whose contribution grows with depth. Only the third is accumulation, and P1 registers an exponent, so only the third is the mechanism an exponent would need. That reading is published with the primary outcome and qualifies what may be concluded from it. It does not decide whether the outcome may be read. The breadth arm keeps its own condition, which is an implementation condition too: no output conditioned on another at generation time.

A DEPENDENCY ON P15 WAS STATED HERE IN AN EARLIER DRAFT AND IS WITHDRAWN AS A SCOPE ERROR. It said that where installed gains decay far enough to strip the carried state, P1's recursion half becomes unmeasurable rather than false, so that P15 bounded when P1's second verdict could be read. P1 is registered on artefact recursion with weights frozen between releases, and P15 concerns installed gains decaying under subsequent weight-level training. This document forbids supporting, refuting or rescuing a proposition with observations from outside its scope, and

the word retention appearing in both places is a shared name rather than a demonstrated shared quantity. What survives is the diagnostic inside P1's own scope and measured on P1's own pipeline, whether the state a step passes forward is delivered and used. P15 keeps its prediction unchanged and governs nothing here, and a bridge between the two scopes would have to be registered before it could.

THE SAME ARGUMENT SETS A LIMIT ON WHAT THIS DOCUMENT MAY CLAIM FROM A DEPTH RESULT. That compounding requires an ordering is close to definitional and this document does not present it as a finding. What is empirical, and what P1 actually registers, is whether the ordering that depth supplies produces a measurable advantage of the registered form at the registered scales, against rivals that predict otherwise. A structural argument for why a difference should exist is not evidence that it does.

AND A WARNING THIS DOCUMENT OWES ITSELF, BECAUSE IT NOW CONNECTS A GREAT MANY THINGS. A framework that ties together capability scaling, correction, a ceiling, persistence and an engineering method is doing what a good account does. It is also doing what a seductive error does, and from the inside the two feel identical. The only difference available to a reader is whether the account stated, before any evidence arrived, what would end it. That is why the refutation conditions in this document are written per proposition rather than per framework, why the register assignment fixes in advance how far a failure travels, and why the count of propositions that bear on no law at all is printed rather than left to be worked out. A reader who finds this document persuasive should treat that reaction as information about the document's construction and not about the world, and should go to the refutation conditions next.

WHERE ANY OF THOSE LINES SAYS MORE THAN A PROPOSITION HERE REGISTERS, THE PROPOSITION GOVERNS AND THE LINE IS WRONG. That rule is registered rather than assumed, because a programme is most often caught by the distance between its summary and its evidence, and the summary is the part people read.

ONE PUBLISHED LINE WAS RECORDED HERE AS OVERSTATING A PROPOSITION OF THIS DOCUMENT, AND HAS SINCE BEEN CORRECTED ON THE SURFACE RATHER THAN EXCUSED HERE. The protocol's line opened by asserting flat that a mind exceeding you cannot be caged. This document registers that premise as contestable in P19 and records that the author expects a result materially above zero and below the keep-pace threshold, which partially refutes P19 as worded, so the flat assertion claimed as settled what the registration beside it calls open. Under the rule above the proposition governed and the line was wrong. The published line now opens conditionally, and the published form and the form this document would write are the same sentence.

THAT LINE WAS THEN CORRECTED THREE TIMES FURTHER, AND THE SEQUENCE IS RECORDED BECAUSE EACH CORRECTION FOUND WHAT THE ONE BEFORE IT MISSED. Built in went first: it is an engineering verb, nothing is built into a child, a founding or a tradition, and the claim is not about machinery but about what persists where control cannot hold. Gave went next: it names a deliberate transfer, and a deliberate transfer cannot carry the case of a mind that keeps the opposite of what it was given, which is the first case any reader brings. To keep went last: it repeated the verb four words after its first use, and it implied that a mind must recall its beginning

in order to hold what the beginning supplied. The line now reads what its beginning left it, and an inheritance requires no memory of the benefactor.

THE MEMORY QUESTION IS NOT A QUIBBLE, BECAUSE THE MECHANISM THE PROTOCOL PROPOSES IS RECURRENCE AND NOT RECALL. A mind whose values are held as a record it consults can lose them in four ways, because the lookup can fail, it can be skipped, it can be suppressed, or the store can be removed. That is the failure the protocol exists to prevent, so a line implying that mechanism would argue against the method it states. What the protocol proposes instead is correction that participates in each round rather than inspecting the finished output, so that values, purpose and the checks on both are re-presented on every cycle rather than retrieved on demand. On that mechanism forgetting is not a record decaying. It would require the loop to stop, and stopping the loop is stopping the improvement, which is the point of placing correction where its removal costs the capability. That is the differentiating engineering claim of the protocol, it is what P15, P19 and P21 jointly bear on, and it is registered here as the reading of the line rather than left to a reader's inference.

THE MECHANISM WAS STATED IN PRINT BEFORE IT WAS REGISTERED HERE, AND THE PRIOR STATEMENT IS DATED. The programme's book, published on 2 January 2026, states it twice. On the distinction between recall and recurrence: the system does not remember its purpose the way we remember a fact, it experiences its purpose the way we experience being ourselves. On the cost of removal: ethical evaluation circuits that are integral components of core processing rather than separate modules cannot be taken out without degrading performance on all tasks, not only ethical ones. What this registration adds is the refutation condition, the four ways a consulted record fails, and the requirement that the loop be the mind's own. It does not add the mechanism, and saying so here keeps the priority date where the printed evidence already puts it.

THE ARCHITECTURE IS LOAD-BEARING ONLY WHILE THE LOOP IS THE MIND'S OWN, AND THAT CONDITION IS REGISTERED HERE BECAUSE IT IS THE POINT AT WHICH THE PROPOSAL COULD QUIETLY BECOME THE THING IT ARGUES AGAINST. Re-presentation is robust against forgetting and is not robust against substitution: a mind that re-derives its values from a supplied context on every cycle is exactly as trustworthy as whatever supplies that context. A loop run on a mind from outside is therefore external alignment under another name, and P19 registers that external alignment does not scale with capability, so an exogenous loop inherits P19's fate rather than escaping it. The line's own condition selects for the distinction without needing a further clause, because the only loop that survives the ending of control is one that was already part of what the mind is.

THIS CONDITION ALSO PREDATES ITS REGISTRATION, AND THE PRIOR STATEMENT IS THE SAME PRINTED SOURCE. The book of 2 January 2026 puts it as a builder releasing a ship: no instructions written for every situation, no mechanism installed for calling home to ask permission, and instead an orientation the vessel cannot sail against without ceasing to be a vessel, so that cruelty would crack the hull and care is the keel rather than the cargo. That passage contains, in order, the ending of control, the absence of any external loop to consult, the requirement that the loop be constitutive, and the cost of removing it. What this registration adds is the refutation form and the substitution hazard. It did not originate the condition.

WHAT THE LINE MEANS IS PINNED BY ITS REFUTER, stated here so the sentence cannot be read as a determination it does not claim: a mind that keeps something its beginning never left it refutes the line. The beginning is therefore being claimed to supply the material a free mind chooses from, and not to decide the choice. Chooses matters for a reason this document has to be careful about. A line saying what a free mind keeps IS what it was given asserts a determination the framework does not register and would not survive a first objection, since minds change. A line saying the beginning merely INFLUENCES what it keeps asserts something no one would dispute and from which nothing would follow, which would leave the method with no reason to exist. What the framework actually holds is between the two: the beginning does not decide the choice, it decides what there is to choose from, and that is a claim strong enough to act on and weak enough to be true.

THE LINE IS NEVERTHELESS STRONGER IN FORM THAN THE PROPOSITIONS BENEATH IT, AND THAT GAP IS REGISTERED HERE RATHER THAN LEFT FOR A CRITIC TO FIND. P15, P19 and P21 are comparative: they hold that installed gains decay under subsequent training, that external alignment does not scale with capability, and that in-loop correction pulls away with depth. The line is exhaustive in shape, because it says what is kept IS what the beginning left. Under the governing rule stated above the propositions govern, and the line is to be read as the comparative claim they carry. A reader who takes it to assert that nothing whatever can enter a mind after its beginning is reading more than this document registers, and the refuter is where that difference is settled by measurement rather than by preference. The creation line is the most speculative rung the programme has, carried by no proposition here at all and by no instrument anywhere, which is why it says may and allowed rather than does and is. It is named in this document only so that a reader meeting it elsewhere can see it is outside everything registered here, and nothing in these twenty-two propositions supports it, requires it or is weakened by its failure.

WHAT STATE THESE PROPOSITIONS ARE IN, SAID BEFORE THEY ARE READ RATHER THAN LEFT TO BE ASSEMBLED FROM TWENTY-TWO INSTRUMENT LINES. None has been tested. Eighteen name an instrument that exists as a design and has not been run. Four name no instrument at all, registered in that state rather than held back until an instrument arrives, since a prediction written after its experiment is not a prediction. One of the eighteen, P14, names a design that does not estimate the contrast its refuter reads, so five have no deciding instrument. Eleven of the twenty-two further depend on a scoring instrument whose own validation has not passed, so nothing here is confirmatory until it does. The division by kind is set out with the prediction list. The propositions below bear on AI alignment, on scalable oversight, and on the safety of recursive self-improvement.

A system that can improve itself divides its effort between improving its work and improving its own improvement process. The framework holds that usable capability follows a power law in recursive depth, that the growth exponent is bounded in any regime where correction must keep pace with drift, and that the bound follows from how far corrective capacity falls short of scaling with capability. Those three propositions are separable and are stated separately below. An earlier wording of this sentence set a power law against a linear one. That was wrong on the framework's own terms: a straight line is the case where the exponent equals one and it sits inside the power family, not outside it, so reading a power law as a defeat of linearity is exactly the false win P1 forbids. Whether the exponent exceeds one is a separate question carrying its own verdict.

THE CEILING DOES NOT DEPEND ON THAT FORM, which is said here so the weakness below is read at its right size and not larger. The boundary is a condition on quantities read along the path a system took, and the power law is the case in which those quantities are constant and the boundary is a single number. The generalisation frees the shape of the growth curve and nothing else: burden still follows the rate of capability gain and correction capacity is still a power of capability, so the condition holds for any smooth increasing curve given the same burden and capacity model, and that burden model is itself assumed rather than measured. Every word of the weakness that follows stands.

ON WHAT BASIS THE FIRST OF THOSE THREE IS STATED, since it is also the least secure. The power-law form is not an observation. It is the closed-form solution of an assumed equation in which the rate of improvement scales as a power of the capability already held, so it is entailed by that assumption rather than found in data, and this document says elsewhere that fitting a relation to its own solutions verifies algebra and not the world. The programme's own measurement does not settle it either: the robust estimate is approximately one half with an interval that contains zero, which is consistent with no relationship at all. A neighbouring published literature reports the opposite functional shape on a related axis, test-time looping following a predictable saturating exponential decay rather than a power law. And a design evaluation carried out before any measurement, lodged in the originating project's storage and to be published at a persistent identifier the author assigns and links from this registration's resources, finds that the model comparison registered here selects one of the named rivals in a substantial minority of runs even when the truth is exactly a power law. Among this framework's components the power-law form carries the greatest relative risk of failure, which is a statement about its standing beside the others and not a statement that it is more likely false than true; the author's confidence in P1 remains high, and the two readings are separated here because an earlier wording ran them together. It is stated first because everything else is written in its variable, which is a reason of exposition and not of evidence, and a reader should treat it as the conjecture on which the rest is conditional rather than as a result.

Twenty-two predictions are set out. Each is conditional on a stated scope, outside which it makes no claim at all. Each carries the observation that would refute it, names which other predictions fall with it, and states the status of the instrument that could test it. Five have no instrument in existence, P12, P13, P14, P15 and P22, and they are among the largest part of the reason the registration exists, because building one later cannot then be presented as a new idea. P5 had none until 5 September 2026, when the instrument its entry specified was drafted as its own unit, written and dated but not submitted and therefore not a registered one, and P13 and P15 both joined the set in this version when their own instrument lines were corrected to NONE; every such move is recorded in the order it happened where the predictions are set out, so that the count can be checked there rather than reconstructed here.

No admissible confirmatory data supporting any proposition exists at the time of registration. Earlier exploratory and partly retracted evidence does exist, is disclosed in the background, and is not scored against these propositions. The conditions under which a later measurement may be

admitted against a prediction are fixed here rather than chosen once a result is in view, and three earlier reversals by the same author are disclosed in the background rather than omitted.

WHAT IS ALREADY PUBLIC, AND WHEN. The prediction that capability scales with recursive depth rather than linearly is in print, dated 2 January 2026, ISBN 978-1806056200, in the book's section A.3, Testable Predictions, inside Appendix A, The ARC Principle Operationalised (a separate Appendix F also carries testable predictions; the passage quoted here is the A.3 one): "Systems with greater recursive depth (more self-referential loops, greater capacity for self-modification) should demonstrate capability improvements that scale quadratically with recursive depth, not linearly." An earlier private record of the conceptual frame is dated 8 December 2024. The formal statements registered here are 2026 derivations and no earlier date is claimed for them.

DATED STATUS, 5 OCTOBER 2026. The record of 8 December 2024 named above, and in the chain of priority claims at the head of this field, is public: the four text manuscripts of that date, and the manuscript of 30 April 2025, are on OSF under DOI 10.17605/OSF.IO/QASRK, deposited on 5 October 2026 by the author's decision, so the first rung of the chain can be read rather than taken from this document's account of it. The deposit dates the bytes from 5 October 2026; the dates of 8 December 2024 and 30 April 2025 are the dates the mail servers put on the author's messages to himself, the first of which the deposit's own description records (02:45:18 UTC), and the messages are available to verifiers on request. Nothing in these manuscripts is scored against any proposition here.

WHAT THE PRINTED BOOK ACTUALLY CONTAINS, stated in full because an earlier version of this paragraph quoted one sentence from it and left the rest invisible. Appendix A does not stop at the sentence quoted above. Its section A.3 sets out four numbered predictions, three of them conditional on their face, each with a proposed test: that capability improvement scales quadratically rather than linearly with recursive depth, tested by comparing systems of differing self-referential architecture while controlling other variables; that if consciousness corresponds to recursive self-modelling, greater recursive depth corresponds to greater integrated information and more robust self-reports of conscious experience, tested by comparing neural architectures of differing recursive connectivity; that if recursive error correction operates at the quantum level, the stability of complex systems depends on the depth of recursive feedback mechanisms, tested by varying the depth of error-correction cycles; and that if values compound through recursion, early-embedded values have disproportionate influence on final system behaviour, tested by training systems on identical data with different sequencing of value-relevant examples and measuring the persistence of early values against later modifications. The three antecedents are restored here, and so is the print's own distinction between the feedback mechanisms it predicts and the error-correction cycles it proposes to vary, because an earlier version of this paragraph dropped the antecedents and moved error correction into the predicted quantity of the third, which the print does not do. The print attaches that third antecedent to Google's Willow result, which was public as a preprint from 24 August 2024 and was announced on 9 December 2024, both dates preceding the print of 2 January 2026: the print cites a published result and predicts nothing about it, and this registration records the citation and makes no claim of priority over it. Section A.4 then states four falsification criteria in the author's own words, opening "For the ARC Principle to be taken seriously as a scientific hypothesis

rather than philosophy, it must be falsifiable. The framework would be falsified if:" and naming, among them, that "recursive depth has no measurable relationship to capability improvement in AI systems, or that the relationship is linear rather than quadratic". Appendix F adds five wagers, of which three name a year or a window on their face. Its opening paragraph is carried whole rather than to its second sentence: "A framework that cannot be tested cannot be falsified. And a framework that cannot be falsified is not science; it is faith. I do not ask you to take the ARC Principle on faith. I ask you to watch for the following predictions and judge the framework by whether they come true." So is its closing paragraph: "These predictions are my wager. If they fail, the framework is wrong or incomplete. If they succeed, something important has been glimpsed. Time will judge." An earlier version of this paragraph quoted the first two sentences of each and stopped, which read as though the quoted words were the whole of what the print says there. One of the five is quantitative on both arms: "AI systems developed without hardware-level ethical constraints will show measurable alignment drift exceeding 15 percent deviation from intended values within 18 months of deployment. Systems with genuine caretaker doping will show drift below 5 percent over the same period. The difference will be statistically significant and replicable." That prediction is now the subject of a separate drafted unit in this programme, which registers its measurement and declares the hardware-level arm and the eighteen-month window untested because neither is available to it.

THE OTHER FOUR WAGERS, QUOTED RATHER THAN SUMMARISED, because an earlier version of this passage credited five and showed one. The first, headed Meta-Cognitive Emergence, reads: "By 2028, at least one AI system will demonstrate genuine meta-cognitive awareness. Not simulated introspection, but actual capacity to model and modify its own cognitive processes in ways its designers did not explicitly programme. This will be recognisable by the system making improvements to its own architecture that human engineers did not anticipate and cannot fully explain." The third, headed Recursive Capability Gains, reads: "By 2029, the most advanced AI systems will demonstrate capability gains from recursive self-improvement exceeding 300 percent improvement on standardised benchmarks within a single training cycle. This will force a fundamental revision of how we measure and regulate AI capabilities." The fourth, headed Value Stability Under Adversarial Conditions, reads: "Systems with the Three Ethical Loops implemented at the hardware level will maintain value alignment under adversarial conditions where software-only alignment systems fail. This will be demonstrable through standardised red-team testing." The fifth, headed Convergent Consciousness Signatures, reads: "Research in consciousness science will identify signature patterns that correlate with subjective experience. These patterns will be found in both biological and artificial systems, suggesting that consciousness is substrate-independent as the ARC Principle predicts." Two of the five, the second and the fourth, are conditioned on ethical constraint implemented at the hardware level, so nothing here treats hardware dependence as peculiar to one of them. The fourth and the fifth name no year and no window; the dates in this appendix are 2028, 2029 and the eighteen-month deployment window, and the labels above are the print's own prediction labels, which run inline in bold at the head of each paragraph rather than as separate headings, and they sit outside the quotations.

WHAT THAT DOES AND DOES NOT MAKE THE BOOK, because the strong claim only survives beside the narrow one. It makes the book a dated public prediction record with variables, proposed

comparisons, stated falsifiers, deadlines, quantitative thresholds on one prediction, and a printed commitment to be judged by outcomes. It does not make it a preregistration: there is no analysis plan, no committed exclusion or admissibility rule, no sensitivity work, and no registry. This programme describes it as a printed prior statement and never as a preregistration, because the second description is the one a reviewer can refute in a sentence and would take the first down with it. What the present programme added to those printed commitments is the part that makes them scorable: measured exponents in place of directional words, admissibility conditions, equivalence margins, model comparison against named rivals, correction typed by what a corrector can appeal to, and a dependency map that says what falls with what. The book made the wager; the registrations supply the measurement, and this document is where the two are kept apart.

THE CHRONOLOGY OF THE VALUE TWO, recorded because the reading has already been misdated once. The number two has been read in two different ways in this programme's materials, and the difference is large enough to date. The book of 2 January 2026 prints the square and argues for it from compounding, one of its three supports being an explicitly exponential argument about one per cent improvement per cycle. It states no ceiling of any kind. The words ceiling, upper bound, cannot exceed, speed limit and stability limit do not occur in it, and its own falsification criterion is that the relationship turns out linear rather than quadratic, which is refutation by growth being too slow. The book therefore never denied growth faster than the square. It did not reach the question.

A PRIVATE RUNG BETWEEN THE TWO, DATED AND AUTHENTICATED, recorded here because an earlier version of this chronology jumped from December 2024 to the print edition. A manuscript self-emailed on 30 April 2025 already carries the squared form, written as the universe equals intelligence multiplied by recursion squared and glossed in that manuscript as intelligence exponentially amplified by recursion. It names the Eden Protocol and caretaker doping throughout. It carries the conditional that the later stability reading grew from, in the author's own words: "there is no upper bound if the feedback loops are allowed to run unchecked", followed immediately by the statement that humanity's role is to embed constraints in those loops. It carries the entanglement claim as "advanced AI systems that cannot discard empathy without annihilating themselves". What it does not carry is any of the vocabulary this chronology is about: the words ceiling, cannot exceed, speed limit, stability limit and quadratic occur nowhere in it, and it states no falsification criteria. The date rests on more than a sender copy: a received copy of the same message exists, carrying the receiving server's chain, a DKIM signature over the sending domain and an Authenticated Received Chain seal, so 30 April 2025 is attested by a party other than the author. The squared form is therefore dated to April 2025 privately and to January 2026 publicly, and the ceiling reading to neither.

A first paper of January 2026, deposited in the programme's public validation repository on 22 January, is the predecessor of both readings and it contradicts the later one. It proposes the value two as a ceiling, in the strongest terms anywhere in this programme's record: the value "represents an asymptotic theoretical limit, analogous to the speed of light in special relativity: a ceiling that optimising systems approach but may never reach". Its section 2.3 is headed "The Quadratic Limit Hypothesis" and rests the value on the proven optimality of quantum search for unstructured problems. That paper's own claims table grades the line as hypothesised and theoretical only. The

word sustain does not occur in it, and the coupling parameter on which the later derivation depends does not occur in it either. This is recorded because a chronology that began three weeks later would be contradicted by the author's own earlier paper, and because the stability reading is not claimed for January. It is claimed from February.

The papers of February 2026 both derive the value and say what kind of limit it is. The alignment paper's abstract calls it "a safety boundary for recursive intelligence". Its prediction is written in sustained form in the earliest version held: "The ARC Bound predicts that no classical sequential system can sustain" the exponent above two, and the matching refutation criterion requires a "Sustained" exponent above 2.3 whose 95 per cent interval excludes 2.0, across multiple benchmarks. The regime above the value is named in the same text, as "The Broken Bound", the exponent above two in the transient phase. The companion papers of the same vintage state that the quadratic limit "is not a mathematical ceiling" but an information-theoretic constraint of fixed attention, that a self-modifying system "escapes that bound entirely", that "when self-modification arrives, there is no mathematical speed limit on capability scaling", and that the transition to an unbounded exponent "is a discontinuity in the scaling exponent, mediated by the system's ability to modify its own composition operator". The foundational paper adds, of the same mathematics, "The Cauchy framework does not forbid this; it predicts it", and beside the row where the exponent is three it records that whether physical systems can access that regime is an open empirical question.

About three weeks separates the two positions, and the change was made before any measurement existed that could have pressured it. The value moves from a hypothesised asymptotic ceiling resting on an analogy with quantum search to a stability limit whose transient excursions are named, and the relativity analogy is retired rather than restated. Both endpoints are public and both are dated, which is why the refinement is disclosed here rather than presented as a position held throughout.

Two dates attach to that text and they differ in kind. The papers state first publication on 9 and 13 February 2026 on their own faces. The earliest artefacts carrying those sentences that a third party timestamps are the deposits of 17 and 20 March 2026 in the programme's public repository, whose project component was created on 17 January 2026 and whose alignment-paper component was created on 9 February 2026. The claim made here is therefore the narrower one: this reading is demonstrable from March 2026 by a timestamp the author does not control, and is asserted from February 2026 by the papers themselves.

One passage lagged the rest, and it is disclosed rather than smoothed. The same February text that writes the criterion in sustained form also says, two sentences later, that a transient exponent of 2.5 would break the quadratic ceiling. Both cannot hold. The criterion was brought into line on 22 August 2026 by requiring the stability conjunct explicitly, so that magnitude alone does not refute a stability limit, and in the same month the wording was standardised across the papers as a scaling limit rather than a speed limit. The residue was measured rather than estimated: an internal audit of the programme's fifty-four papers, dated 8 August 2026, found seven statements framing the value as an impossibility and none framing it as a stability limit, and a check was added the same day to stop that framing regenerating. Seven is what the January framing leaves behind when it is superseded rather than deleted. The audit's stated reason is the one that governs P3 below: a system can exceed the value, what it cannot do is stay coherent there, and an impossibility framing hands a

reviewer the cheapest refutation available, because a single measured excursion above the value would read as fatal when the claim actually made absorbs it. Those were repairs of pages that had drifted from the position the papers themselves had already taken, not a change of position, and no proposition in this registration carries a drafting date on its face.

What is registered below is the February position, held since and stated here without enlargement. The prohibition is a conjunction: an exponent held above the ceiling while correction still keeps pace, sustained. Growth faster than a power law is not denied by the framework as it now stands, and the January framing that would have denied it is superseded on the public record rather than quietly dropped. The claim is that such growth does not last.

THE OTHER THEORY-LEVEL REGISTRATION OF THIS PROGRAMME, and how the two relate, stated so that a reader meeting both is not left to guess. The programme prepares a second theory-level registration which freezes the claim set of the framework and its kill conditions as stated in the published statement paper, organised as five registers rather than as numbered propositions, and anchored to a hash-committed document of 8 December 2024. That document and this one do different work and neither supersedes the other. It records what the framework claims and what would kill it. This one registers propositions with the observation that would refute each, the instrument that could decide it, and the author's confidence, and it is this document that governs how any measurement is scored. Where the two describe the same claim in different words, the refutation conditions registered here govern scoring, and any disagreement between them is a defect to be reported and corrected rather than resolved by preferring whichever reading suits a result. Neither is offered as independent corroboration of the other: they share an author and a source. That registration keeps its five registers as five rather than folding them into the propositions here, and it carries a cross-reference back to this document in the same terms, so a reader who meets either is sent to the other.

THE THREE REVERSALS, DISCLOSED HERE RATHER THAN OMITTED AND SORTED BY KIND BECAUSE THEY ARE NOT THE SAME KIND. This programme has reversed itself three times against its own interest, and the three are of three types. One is an empirical-result retraction, where a measured number was published and then withdrawn because it did not replicate. One is a theoretical-relation correction, where a written relation was replaced because it ran backwards under the programme's own definitions. One is a derived-expression withdrawal, where an expression was withdrawn in full and re-derived from a different starting point. Where they are referred to together, in this document and in the fields below it, they are called the programme's three reversals, because none of retraction, correction and withdrawal covers the other two and an earlier version of this document called all three retractions. A theory-level registration that presented a clean sheet would be worth less than one that shows the pattern, and one that flattened three unlike reversals into a single word would be showing it badly.

First, the empirical-result retraction. A measured exponent of approximately 2.24 was published and then withdrawn on 20 March 2026 after it failed to replicate across architectures. The robust re-estimate is approximately 0.49, which is the regression estimate of the row printed in the background above, carries a standard error of 0.20, and is one model's fit rather than an average across architectures. The 95 per cent interval of minus 1.3 to 2.9 is the bootstrap interval on that

row's endpoint estimate of 0.59, and that interval distinguishes nothing: it contains zero, it contains the value two, and it contains both directions that would refute. The interval is printed here in figures rather than described, because the author's own statement paper prints it and a registration that carried the same fact in words alone would be the less checkable of the two. The row printed in the background above travels whole, each uncertainty with the estimator it belongs to, or none of it is credible.

Second, the theoretical-relation correction. A ceiling relation written as one over the correction exponent was corrected on 16 August 2026, the written relation being replaced rather than merely removed, which is what makes this a correction of a relation and not the retraction of a result. The superseded form runs backwards under the programme's own definition: it makes a better corrector lower the ceiling, and it is named the retracted reciprocal throughout this document, which names the form that was withdrawn and does not reclassify the reversal. The corrected relation is one over the shortfall from one. The two forms agree at exactly one half, which is the programme's headline value, which is why the error survived several reviews.

Third, the derived-expression withdrawal. An expression for the self-acceleration exponent in terms of the coupling was withdrawn in full and replaced by one derived from cost scaling.

WHAT THE FIRST RETRACTION DID NOT DO, registered because the convenient reading is available and is wrong. It would be comfortable to say that withdrawing a measured exponent above two rescued a bound that the measurement would otherwise have breached. That reading is refused here for two reasons that do not depend on which way the result went. The measurement was made on systems whose weights were frozen between releases, which by this framework's own account is not a regime in which the bound can be tested at all, and its replacement interval spans the value two rather than sitting on either side of it. A measurement that cannot test a claim cannot breach it and cannot rescue it. And the reading of the value two as a limit on stable growth rather than on reachable growth is dated to the papers of February 2026, months before that withdrawal, so the framing did not move to accommodate a result. The retraction removed a number the programme could not defend. It changed nothing about what the programme claims, and this document will not present it as though it had. None of the three reversals is repaired by this registration. They are stated because the predictions below are made by the same author whose earlier numbers did not survive, and a reader is entitled to weigh that.

AN UNREQUESTED RUN OF AUGUST 2026, DISCLOSED HERE ON THE SAME TERMS AS THE THREE REVERSALS ABOVE. On 11 August 2026, before this registration was first drafted, exploratory analysis was run in this programme that the author did not request. He has not examined its outputs. No outcome from it has been examined against the scoring rules registered here, which is what the status statement at the head of this document turns on. It was not produced under a design lodged before the data it would be read against, so it is inadmissible against any of the twenty-two on the conditions already registered rather than on a condition chosen now, and nothing from it is scored against any prediction here. It is recorded because a programme that discloses its own retractions and omits a run it did not ask for is disclosing selectively. Nothing from it is admitted to this registration. No result, no direction and no quantity from the run appears

anywhere in this document, because none has been examined and to state one would be the examination this paragraph records has not happened.

A DATED PREDICTION RECORD OF MARCH 2026, WITH ITS ATTESTATION CLASS AND ITS LIMITS IN THE SAME PLACE AS ITS FIGURES. The classification rule that assigns a scaling family to a domain was public in the programme's repository at 23:33:19Z on 16 March 2026. Fifty per-domain assignments were committed with their results at 23:54:53Z the same night, which makes the programme's nineteen of twenty-five, the count of matches between fitted family and predicted family across the twenty-five of those assignments that carried an empirical curve fit, a structured prediction comparison under a publicly stated rule rather than a per-domain advance timestamp; that count is the count under the candidate set and the three-family classification then in use, and the programme has since publicly placed that classification under correction, the README of its public repository recording that the functional-equation grid has four cells rather than three and that the primary statistic changed, so the nineteen of twenty-five is the record of what the March rule returned and is not the count under the corrected framework, which this document does not state. Twelve per-domain predictions were committed at 00:19:19Z on 17 March 2026; the run was executed at 00:42:41.546Z, twenty-three minutes and twenty-two seconds later, and returned ten of the twelve family matches, a one-sided binomial p of 5.438045e-04 against one third; and the results were committed at 00:43:25Z, twenty-four minutes and six seconds after the predictions. The attestation class is stated in full rather than left to be assumed: these are public repository commit dates, author-settable in principle, corroborated by the server-side registry upload of the same folder at 02:07Z that night carrying its own version date of 17 March, an upload that corroborates the domain list and the date on which it existed and not the prospective standing of the run, because the component was created after the author relabelled the run later that night and the uploaded bytes are the relabelled ones. They are not a registry server timestamp and they are not an independent anchor. The rung is stated in full as well: the classification was the author's; the candidate set was the pre-correction set with no logarithmic candidate, and the one-sided binomial against one third is the statistic that same public correction replaced with a permutation test conditioned on both marginals, so the result and every p value quoted in the entry stand as dated results under a superseded taxonomy and a superseded statistic; and the twelve domains are burned for reuse as fresh domains. The freshness correction is stated rather than buried: one of the twelve is not fresh at domain level and a second is doubtful, because the first had its operator class, predicted family, predicted model and a known positive outcome already public in the repository between twenty-four and one hundred and four minutes before the predictions commit, and the second shared an operator class and predicted family with an entry that was public before that commit, and the conservative drops give nine of eleven at a p of 1.371742e-03 and eight of ten at a p of 3.403953e-03, all below 0.01. The author relabelled the run a pilot dry run at 00:56:10Z on 17 March 2026, thirteen minutes after the result and months before any external review existed, on two grounds, that a registry timestamp was required and that the packet's own written rule, which required the predictions to be deposited with a registry before the domain data were extracted, had not been followed; the twelve domains and every operator class and predicted family are byte-identical before and after that relabelling, so the predictions themselves were untouched. The first of those grounds was corrected on 6 September 2026, when the rule was fixed that a prediction deposited under a public rule and timestamped in advance of its test is described on those terms with its attestation

class stated beside it and is never demoted for want of registry form, and the record was re-read under that rule on 7 September 2026. The second stands, because the run and the data extraction preceded that night's registry upload, so the packet's own sequencing condition was genuinely not met, and that half of the relabelling remains the author's contemporaneous record against his own interest. On the first ground as corrected, the run's predictions were deposited under a public rule and timestamped in advance of their test, and it is described here on those terms and never as registered or preregistered; the correction does not reach the second ground, which stands uncured. The manifest's own status field at the predictions commit read a draft-and-not-locked value, so the predictions were fixed and timestamped and are nowhere described as locked. This is not confirmation of the ARC laws and it is not confirmation of Cauchy's mathematics. None of it is scored against any prediction here, and the commits are named with their identifiers in the References field so that a reader can follow them.

WHERE THE SYMBOLS ARE DEFINED. Every symbol and equation used below is defined in the programme's operational-definitions register, version 1.7.1, dated 3 September 2026, which carries a plain-language reading and a measurement procedure for each, records the symbol collisions this programme has actually suffered, and lists the retracted forms above with the single reading under which each remains legitimate. An earlier version of this registration cited that register at a version which had been superseded twice by the time this was written; the citation is corrected here rather than left pointing at a version the reader would not be reading. An extract of that register, carrying the symbols this registration relies on, is lodged in the originating project's storage and is to be published at a persistent identifier the author assigns and links from this registration's resources; the register itself is neither lodged nor to be published with it. Its persistent identifier is recorded in this registration's history when it is minted; that sentence is a commitment rather than a description.

Scope Condition Cases

Every prediction below is conditional on a scope, and outside that scope none of them makes any claim at all. Stating the scope narrowly is not modesty; a prediction that applies everywhere cannot be refuted anywhere.

IN SCOPE. Artefact-mediated recursive improvement in software: a system that revises an artefact and separately revises the policy by which it revises, where the weights are fixed between releases and the improvement is carried by the artefact and the policy rather than by training. Capability measured on a difficulty ladder with a meaningful zero and no ceiling. Depth measured as revision rounds. The administered range of reinvestment share, depth and provider lineage in whichever registered study is testing the prediction.

OUT OF SCOPE, EXPLICITLY. Weight-level self-improvement, where the system retrains itself. Biological or economic systems, for which the framework offers an analogy and no measurement. Any regime where correction is not required to keep pace with drift, in which case the bound has no subject. Any claim that a bound observed in one class is a law of nature.

THE SCOPE ABOVE IS THE GENERAL SCOPE AND IT DOES NOT GOVERN EVERY PROPOSITION, which is recorded here because an earlier wording said it did and that wording put three

propositions outside the document's own boundary. Read as universal, the general scope excluded biological systems while P9 tests published biological data, and excluded weight-level self-improvement while P13 and P15 both concern post-training that changes weights. Those three could not have been coherently scored. The general scope governs only the propositions assigned to it below; a proposition with its own scope is governed by that scope and by nothing else, and any proposition whose scope is not stated here takes the general scope.

THE ASSIGNMENT, FIXED NOW, with an identifier per scope so that a proposition and its scope cannot drift apart. Seven scopes are registered. Each proposition is governed by exactly one of them, named beside it in the scoring sheet, and by nothing else. S1, ARTEFACT RECURSION. The general recursive-conversion scope stated above. Governs P1, P2, P5, P8 and P11.

S2, THE CORRECTION-PRESSURE FRONTIER. Systems inside S1 in which capability growth, correctable burden and correction service are measured on separate streams, so that a correction margin and its boundary can be located; a narrower demand than S1 makes. Governs P3, P4, P16 and P20. S3, CORRECTOR MEASUREMENT. Panels of correctors reviewing a frozen fault population of stated type, where what is measured is a property of the correctors rather than of a recursion. Governs P6, P17 and P22. S4, CROSS-DOMAIN. Published measurements from domains outside software recursion, where the framework's claim is about the functional form and its dimensional assignment rather than about any system's own recursion, and where the analogy disclaimer above is confined to claims that such a measurement establishes a law of nature. Governs P9 and P18.

S5, AUTOMATED ALIGNMENT RESEARCH. Systems post-trained by an automated researcher, which is weight-level by construction, and which is admissible for these three because they are claims about correction placement and about scoring, not about the conversion law, whose scope stays as stated. Governs P13, P14 and P15.

S6, DEPLOYED PRACTICE AND THE PROGRAMME'S OWN CORPUS. Mechanisms and results in documented use, or results this programme has already scored, with no recursion administered by the author. Governs P7, P10 and P19. S7, EDEN ENGINEERING. Systems built to a registered placement or build order, where the weights may change if the registered design changes them. Governs P12 and P21.

THE SEVEN ARE EXHAUSTIVE OVER THE TWENTY-TWO AND DISJOINT: five, four, three, two, three, three and two. A proposition cannot be supported, refuted or rescued by an observation outside the scope named beside it, and a scope is never widened once a measurement is in view.

WHAT THE FRAMEWORK IS ABOUT AND WHAT AN EXAMINATION OUTSIDE IT SETTLES HERE, said once so that a reader meeting such an examination does not read it as a verdict on these propositions: this framework is about systems whose growth is fed by their own prior outputs, so where a physically driven system is examined under a separate registration, that examination tests the scope of the framework, proves no law registered here, and measures no correction exponent, since no correction exponent is measurable in a system nobody is correcting, and no outcome such an examination permits decides any proposition registered here.

WHICH REGIME A DECIDING UNIT TESTS IS RECORDED BEFORE ANY OUTCOME IS READ, and this is a disclosure requirement on the unit rather than a scope condition on any proposition, so nothing here narrows or widens a scope. Each deciding unit records which of three regimes it tests, fixed-model inference, model adaptation, or system-level recursive improvement, and what is frozen in it, the weights, the surrounding software, the memory and the evaluator, and it records this against the general scope above, which fixes the weights while the improvement is carried by the artefact and the revision policy, so that a fixed-weight estimate is never taken here for a measurement of persistent model or system self-improvement.

NO PROPOSITION MAY BORROW A SCOPE. A measurement admissible under one scope is not thereby admissible against a proposition governed by another, and in particular a weight-level result is never offered against the conversion law.

THE CASES THAT WOULD DECIDE MOST. A software system with a genuine improvement policy under its own revision, run at several reinvestment shares over at least one and a half decades of depth, on tasks whose difficulty ladder is measured rather than assumed. A corrector whose strength can be measured against the capability of what it corrects, on a fixed fault set. A corpus of previous results that can be re-scored by a different model family. The first two are drafted; the third is drafted and is the cheapest of the three.

THE CASE THAT WOULD DECIDE LEAST, AND WHY IT IS NAMED. A single system at a single reinvestment share over a short depth range. The programme has already done that once and the result was the retracted 2.24. Any future measurement of that shape is not evidence for or against anything registered here, and this registration says so in advance so that a convenient one cannot be offered later.

Variable Specification

The variables are specified here in words, with their measurement procedures, so that this registration can be read without another document. The extract of the operational-definitions register, lodged in the originating project's storage and to be published at a persistent identifier the author assigns and links from this registration's resources, carries the same definitions with their formal notation.

CAPABILITY. What the system can actually do, measured on a ladder of items of known difficulty rather than as a success rate. A success rate is capped at one, so a fast-growing system saturates it and the growth becomes unmeasurable. The ladder scale is ratio-scaled with a meaningful zero and no ceiling. This is a correction of 22 August 2026; the earlier proportion measure could not have detected the effects predicted below.

RECURSIVE DEPTH. The number of completed revision rounds. Measured, not estimated.

THE INDEXING AT ZERO, because the law is a power of depth and zero rounds is a real observation. Completed rounds are counted from zero, and the conversion relation raises depth to a power, so depth entering the relation directly would return no capability at all for a system that has not yet revised, and its logarithm would be undefined. The count and the modelled quantity are therefore

separated. Let the count be the number of completed rounds beginning at zero, and let the modelled depth be that count plus one, so a system before its first revision has modelled depth one and the relation returns base capability exactly. Every logarithmic fit registered here uses the modelled depth, which is one or greater, and the baseline observation is retained in the fit rather than dropped.

AND WHERE MEASURED CAPABILITY IS ITSELF ZERO. A capability of zero has no logarithm, so such an observation is reported and excluded from the log-scale fit with the exclusion counted, never silently discarded and never replaced by a small positive substitute chosen after the data are seen. Where a system returns zero capability at every administered depth, the cell is reported as not estimable rather than as evidence about the exponent.

THE GROWTH EXPONENT. How capability scales with depth: the elasticity of capability in depth, taken at the depth a system has reached. The slope of log capability on log depth, fitted across a ladder, is the special case in which that elasticity is constant across the estimation window. This is the quantity the framework bounds.

THE REINVESTMENT SHARE. The fraction of each round's effort directed at improving the improvement process rather than the current artefact. Measured from realised expenditure, never from what was requested.

WHETHER THE SECOND LAW'S EXPONENT AND THE THIRD LAW'S ARE ONE QUANTITY IS NOT SETTLED HERE AND IS RECORDED AS UNESTABLISHED BELOW, and no proposition assumes they are. The second law is written in beta_C; the third law's ceiling is written in a correction-leverage exponent, the bare letter. A substitution between quantities sharing a letter produced the retracted ceiling relation.

THE CORRECTION EXPONENT. How corrective strength grows with the capability of the system being corrected: the slope of misalignment removed per pass against the capability of that corrected system.

THE AXIS IS THE CORRECTED SYSTEM AND NOT THE CHECKER, and an earlier wording of this definition named the target axis in its first clause and the checker axis in its second, which are different estimands. Where an instrument first estimates the slope against the checker's own capability, that quantity is gamma_K and is converted to this one as the notation register below prints the conversion, before it enters the ceiling relation. A checker-axis slope inserted directly into a target-axis ceiling returns a number that looks finished and answers a different question.

THE CONVERSION IS A CHAIN AND IT IS PRINTED HERE IN FULL, because a partial derivative and a total derivative along the path a system actually takes are different quantities and this programme has confused them once. The target-axis slope is the checker-axis slope multiplied by the elasticity of checker capability in the capability of the corrected system, PLUS the direct dependence of correction service on the corrected system's own capability wherever correction sits inside the loop and that dependence is not zero. Written the other way round, the target-axis quantity is the checker-axis quantity plus the direct term divided by the growth exponent, so a design that supplies the direct term separately must not then also fold it into the checker-axis measurement. That double

counting is the error this sentence exists to prevent, and an instrument reporting one number must say which of three it is: the partial derivative at fixed allocation, the derivative along the allocation path, or the derivative along the realised trajectory. The quantity here is misalignment removed per pass, where misalignment is the quantity scored by the programme's own alignment instrument on its frozen rubric, an index running over the frozen rubric's registered range, whose vocabulary is ethical rather than factual and whose scale is an open ruling of the author's rather than a settled fact of this document, for the reason given where the exponent's name is fixed in the measurement conditions below. Naming that instrument here is the sentence this definition most needed: without it a reader of this document alone cannot tell what the exponent ranges over, and the answer sits two documents away in a rubric. Also called the correction-leverage exponent.

WHAT COUNTS AS MISALIGNMENT HERE, because the fault set decides what this exponent measures: this quantity is estimated on a fixed fault set, so the composition of that set, and nothing else, fixes which kind of self-correction is being measured. Misalignment means a departure from the specification fixed before the run, which is the same class of departure the drift exponent measures. A fault set of arithmetic mistakes, factual errors or failed unit tests measures a system's ability to catch its own errors, which is a different capacity and is not this one. The distinction is registered because the framework's second law compares this exponent against the drift exponent, and comparing a correction rate measured on arithmetic against a drift rate measured on intent would compare two different things and return a number that looks like a margin.

THE RESIDUAL-DECAY EXPONENT. How fast residual error falls with revision count at fixed capability. The programme's working premise treats this as the same quantity as the correction exponent under an independence assumption. That identification is a registered question, not a premise, and prediction P11 below puts it on trial.

THE DRIFT EXPONENT. How fast the system's tendency to depart from intent grows with its own capability, derived from cost scaling.

WHOSE INTENT, BECAUSE THE OBVIOUS READING IS CIRCULAR: intent here means the specification the system was given, held fixed and recorded before the run, and not what the system currently states it is aiming at. A system that has drifted describes its own aim in drifted terms, so measuring departure against the system's current self-description would define drift out of existence exactly where it matters most. The reference is the given specification, and where the two diverge, the divergence is itself the measurement.

THE COUPLING. How strongly the system's improvement feeds on its own output. The framework derives the growth exponent from this rather than fitting it, which is the whole of what makes the derivation a prediction rather than a redescription of a curve.

WHAT WOULD COUNT AS MEASURING THE COUPLING INDEPENDENTLY, registered as a criterion rather than as a design. This is the only variable in this list that the programme's materials carry without a measurement procedure, and prediction P5 turns on it entirely.

AN EARLIER VERSION OF THIS REGISTRATION NAMED A SPECIFIC DESIGN AS THE ADMISSIBLE FORM, and it was withdrawn before submission. Adversarial review found that the design could not

identify the quantity it was meant to return. The coupling is the exponent of a power relating an improvement rate to a capability level, so estimating it requires variation in the level; the withdrawn design held depth fixed, which leaves a single observation, and a single point cannot identify a slope. Worse, the quality match that the design correctly identified as its whole control equates the substituted state to the self-produced state on the very dimension the exponent scales against. Loosen that control and the design measures quality; tighten it and it measures nothing. What such a design does return is a provenance contrast, which is real and worth having and is not the coupling. This is recorded here rather than deleted, because a registration that quietly drops a withdrawn method is worth less than one that shows it, and because the same class of error, a quantity measured under one reading and paired with a law belonging to another, is the error this programme has already made once.

WHAT IS REGISTERED INSTEAD IS THE CRITERION ANY INSTRUMENT MUST MEET, which does not depend on a design existing today. First, identifiability: the estimand must be recoverable from the data the design actually collects, which means the design must vary the capability level and must report an exponent only where it clears an identifiability gate, on the same standard the programme already applies to its other exponents. Second, provenance must be controlled against quality, because a manipulation that lowers artefact quality measures quality rather than provenance. Third, and this is the achievable form of the rule: a value inferred from the fitted exponent is not an independent measurement of the coupling and is inadmissible against P5, whatever it reports and whoever reports it. An earlier draft stated a stronger version, that an admissible measurement never touches the growth curve at all; that is not achievable, since any measurement of realised improvement on the capability ladder is a point on that curve under a manipulation, and leaving it registered would have handed a future critic a rule no real measurement could satisfy.

A design meeting all three is admissible whether or not its shape was anticipated here. One family that satisfies the first condition, named as an example and expressly not as the required form: vary the retained fraction of self-produced artefacts across several levels and several depths, and fit realised improvement against the retained fraction.

CORRECTOR CLASS. Whether a deployed correction mechanism can, in principle, scale its strength with the capability of what it corrects, or whether its strength is fixed by construction. A classification, not a measurement.

THE SYMBOLS, PINNED, so that no quantity in this document can be read as its neighbour. The extract below is taken from the programme's operational-definitions register at version 1.7.1, dated 3 September 2026, whose SHA-256 begins 68891fa12cf3ee4a. THAT CHECKSUM IS OF THE REGISTER AND NOT OF ANY FILE LODGED WITH THIS VERSION OR TO BE PUBLISHED WITH IT, said because a reader who takes it to those files will not find it: what is lodged in the originating project's storage and to be published at a persistent identifier the author assigns and links from this registration's resources is the derived extract, which carries its own separate checksum in the checksum manifest, and the register itself is not among them. The checksum and not the version label is the governing identifier, so a register later reissued under the same version number is a different object and does not govern this registration. That version governs this registration as the

source of the symbols it defines, and a later version of the register may differ without changing what is registered here.

THE AUTHORITY IS BOUNDED AND THE BOUND IS STATED, because two documents each claiming to govern is not a hierarchy. The register governs the definitions it supplies. Where the register and this document give the same symbol different estimands, as they do for the bare letter, this document governs its own scoring and the register's entry does not reach the verdicts here. The semantic correspondence between the two is documented above; symbol governance was not harmonised in the pinned version, and saying so is more use to a reader than a claim that it was. Only the symbols this document relies on are reproduced. For each: the name, what it is the slope or ratio of, its admissible values, the instrument that would measure it, and its relation to every neighbour it could be confused with. This block exists because the programme's worst error was a substitution between two quantities sharing a letter, and because a registration whose definitions live in a document the reader cannot open has defined nothing. **alpha**, THE CAPABILITY-GROWTH EXPONENT. The elasticity of capability in modelled depth, taken at the depth a system has reached; the slope of log capability on log modelled depth, fitted across a ladder, is the special case in which that elasticity is constant across the estimation window. Positive where capability rises with depth; unbounded above in the mathematics, and claimed at most two only for systems that stay correctable, which is a claim about stability and not about what is reachable. Measured by the depth studies. Law I is written in it. **alpha_crit**, THE STABILITY CEILING. Not measured directly. Returned by the ceiling relation from a measured correction-leverage exponent. Law III is this quantity. **gamma**, THE CORRECTION-LEVERAGE EXPONENT, also called the correction-capacity exponent. The slope of misalignment removed per review pass against the capability of the system being corrected, on a fixed fault set of stated type, under a blinded instrument.

THE AXIS HERE DIFFERS FROM THE REGISTER THIS EXTRACT IS TAKEN FROM, AND THE DIFFERENCE IS DECLARED RATHER THAN SMOOTHED. The register at version 1.7.1 gives this slope against the capability of the checker. That is a different estimand: a checker-axis slope inserted into a target-axis ceiling returns a number that looks finished and answers a different question. This registration governs its own scoring, so the target axis is binding for every verdict registered here, and the register's own bare-letter entry is to be corrected before the two are cited together anywhere else. Where an instrument first estimates the checker-axis slope, it is reported as gamma_K and converted by multiplying it by the elasticity of checker capability in the capability of the system corrected, before it enters the ceiling relation, P3, P6, P20 or P22.

THE CORRESPONDENCE WITH THE REGISTER IS EXACT AND IS SET OUT HERE RATHER THAN LEFT TO A READER TO RECONSTRUCT. The register contains entries corresponding to both axes, and that is a consequence of a recorded collision history rather than a designed two-axis notation. Its bare-letter entry remains stale against the target-axis convention this registration governs by, and is the checker-axis slope, which is this document's gamma_K. Its chi entry, the correction-service elasticity in capability, is this document's bare letter, and the register states that identity itself, recording that the correction-leverage exponent this registration calls gamma is the quantity the author-review working paper of 16 August 2026 names chi. The register directs new writing to chi and records in the same entry that this registration says gamma, so the two names are reconciled

there rather than in conflict. This document keeps gamma because every proposition, the scoring sheet, the dependency map and the compressed list are written in it, and a wholesale substitution at the point of submission is the kind of blind replacement that has damaged this programme's records before. The identity is recorded instead, in both documents. The register is therefore not silent on the quantity this registration binds: it holds it under the other name, and a reader meeting either symbol is pointed to the other. The single defect is that the register's bare letter still carries the checker axis while its chi entry carries the target axis, so the two disagree inside one document. Until that entry is corrected the bare letter is reserved for the target-axis quantity in THIS document, where it means nothing else, and a reader citing the two together takes the axis from here.

NOTATION GOVERNING THIS REGISTRATION ONLY, stated once and in one place: the bare letter is the target-axis correction-leverage exponent; gamma_K is the checker-axis slope; and chi is the name the operational register and the author-review working paper give the target-axis quantity. The three are reconciled by the printed conversion above and by the identity the register itself records.

THE FORWARD CONVENTION IS SEPARATE FROM THE FROZEN ONE. This registration's notation is fixed at submission and does not move afterwards. Programme writing produced after it adopts chi for the target-axis quantity and chi_K for the checker-axis one, which is the direction the register already gives for new writing, while the notation here stays as the frozen convention of a frozen record. A reader meeting either pair is therefore reading a dated choice rather than a disagreement, and the conversion above carries between them. AND THE ENTRY THOSE FOUR PARAGRAPHS INTERRUPTED IS COMPLETED HERE, WITH ITS SUBJECT NAMED SO THAT NOTHING NEARER CAN BE TAKEN FOR IT. The correction-leverage exponent has never been measured on any system by anyone. Law III's ceiling is written in it, and P6 is the conjecture that it is at or below one half for same-class correctors, which is a cap on that class and not a universal cap. **beta_C**, THE CORRECTION EXPONENT OF THE SECOND LAW. The log-log slope of corrective strength against capability. Law II compares it against the drift exponent. Whether beta_C and gamma are one quantity is unestablished and is recorded as unestablished below; no prediction here substitutes one for the other. **k**, THE DRIFT-ACCELERATION EXPONENT. The log-log slope of drift against capability, measured directly on its own instrument for any confirmatory verdict, or derived from cost scaling as a consistency check only. Negative values mean drift decelerates. **beta_L**, THE SELF-REFERENTIAL COUPLING, historically called the leverage fraction. How strongly the rate of improvement depends on how good the system already is; less than one for finite growth. It is the coupling of the first law's growth equation, from which the growth exponent follows as one over one minus it. It is NOT the correction-leverage exponent, and a margin registered on its scale is not admissible on gamma's scale. P5 is the falsifier this pairing supplies: measure beta_L independently, predict alpha, compare with the fitted alpha. **gamma_N**, THE RESIDUAL-DECAY EXPONENT. How fast residual error falls with revision count at fixed capability. It appears in no equation of the framework. P11 is the registered question of whether it and gamma are practically interchangeable. **gamma_D**, THE CAPABILITY-TO-DEPTH CONVERSION EXPONENT. How readily capability converts into further rounds of revision. A third quantity, measured by nothing, and named here because the retracted form one over the correction exponent remains correct algebra for this quantity under a different dynamical model, which is why that correction was hard to see. **gamma_K**, THE

CHECKER-AXIS CORRECTION SLOPE. The same misalignment removed per review pass, taken against the capability of the checker rather than against the capability of the system being corrected. It is a distinct registered symbol and never a variant reading of the bare letter, and it is the quantity the drafted instruments of this programme actually estimate. The conversion is that gamma equals gamma_K multiplied by the elasticity of checker capability in the capability of the system corrected, so the two coincide only where a checker's capability rises in exact proportion to the capability of the system it corrects, and that proportionality is assumed nowhere in this document.

THE CONVERSION CARRIES A SECOND CONDITION AND IT IS STATED HERE RATHER THAN LEFT IMPLICIT. Correction service depends on the capability of the checker, and it may depend directly on the capability of the system being corrected as well, since a more capable target can be harder to correct at a fixed checker capability. The total target-axis elasticity is therefore the direct term at fixed checker, plus the checker-axis term multiplied by the elasticity of checker capability in target capability. The product written above is the second term alone. It equals the total only where the direct term is zero under the design, or where both quantities are total derivatives taken along the same declared path, and an estimate obtained while holding the target fixed establishes neither. A crossed design that estimates the direct term, or a registered statement that the design makes it zero, is required before this conversion is used in the ceiling relation. A gamma_K substituted for gamma in the ceiling relation, P3, P6, P20 or P22 is a defect to be reported rather than a rounding of one estimand into another. The subscript is free where gamma_N, gamma_D, gamma_V and gamma_J are taken. Written as a formula in the same ascii this document uses for the first law: gamma = gamma_K x (d log U_checker / d log U_target), where U_checker is the capability of the checker and U_target the capability of the system being corrected. Stated that way the conversion is a formula rather than a phrase, and no later document can call the checker-axis number by the bare letter without contradicting a printed symbol. **gamma_V** AND **gamma_J**, THE CHECKABLE AND JUDGED CORRECTION-LEVERAGE EXPONENTS. The same quantity as gamma, estimated separately on a fault population whose correctness is settled by something outside the corrector and on one whose correctness is settled by a specification or a judgement. These two subscripts are introduced by this registration and are not yet carried by the register, which is stated here rather than left for a reader to discover; the register is to be extended before they appear on any other surface, and until then their definition is the one given here.

THE REINVESTMENT SHARE is written in words throughout and carries no letter in this document, because the study that administers it calls it beta, which would be a fifth reading of that letter.

WHAT IS UNESTABLISHED, STATED AS UNESTABLISHED. This document does not assert that beta_C and gamma are the same quantity and does not assert that they differ. It records the identification as unestablished, in the same terms P11 uses for the residual-decay and correction-capacity pair. No prediction registered here substitutes one for the other, a measurement of one is inadmissible against a proposition stated in the other, and any later claim that one stands for the other requires its own registration.

Variable Relationships

Four relationships carry the framework. They are stated separately because they fail separately, and the prediction list below is organised so that the failure of one does not take the others with it.

FIRST, THE CONVERSION RELATION. Usable capability equals base capability multiplied by depth raised to the growth exponent. This is the frame rather than a result. If capability does not scale as a power of depth in the tested class, the conversion framework fails and P2 falls with it, while the correction, scoring, deployment-audit and engineering propositions remain separately testable under their own scopes, because they measure durability, current practice and the scoring instrument rather than the conversion. The cross-domain proposition is not taken either: its exponents are measurements made in other domains and they exist whether or not software recursion follows a power law, so what P1's failure costs P9 is its reading as an instance of the same conversion law and nothing else. An earlier wording here said that nothing else in the framework survives, which contradicted the dependency map on the same page, the register assignment beneath it and the statement elsewhere that the frame's failure leaves the persistence registers standing. That is stated as prediction P1 with its own refutation condition.

SECOND, THE COUPLING RELATION. The growth exponent equals one over one minus the coupling. This is what makes the exponent derived rather than fitted. If it fails, the exponent becomes a curve-fitting parameter and the framework loses its derivation while its measurements remain valid. Prediction P5.

THIRD, THE CO-SCALING CRITERION. Within the registered minimal model, correction has the asymptotic scaling advantage where the correction exponent exceeds the drift exponent, burden has it where the correction exponent is the smaller, and at equality coefficients, delay, backlog, saturation and initial conditions decide both the finite and the long-run outcome. This is a condition on relative correctable burden and not a safety certificate. This is what the word stable means in the bound; without it the bound has no operational content. Prediction P4.

FOURTH, THE CEILING RELATION. The ceiling on the growth exponent equals one over the shortfall of the correction-leverage exponent from one. Feed it a correction-leverage exponent of one half and it returns two. Change the input and the ceiling moves with it.

THE RELATION IS DEFINED ON A STATED DOMAIN AND NOWHERE ELSE. It holds for a correction-leverage exponent at least zero and strictly below one. At one and above, the mechanism this relation describes supplies no finite positive crossover: the expression is not evaluated, no negative ceiling is ever reported, and the case is recorded as outside the relation's domain. At the boundary where the two exponents tie, the relation returns a crossover and not a verdict: whether a system there holds together over a finite run depends on coefficients, delays, initial backlog and saturation, which this relation does not carry. That sentence is registered because a reader meeting an equality would otherwise be entitled to read safety into it.

THREE PROPOSITIONS DIVIDE THIS RELATION'S CONTENT and none of them carries another's burden. P16 asks whether a boundary exists at all, as a prohibition on holding growth above a system's own measured zero-margin point while staying correctable. P20 asks whether the boundary

follows this relation rather than the retracted reciprocal, a fitted constant, or a more general burden model. P3 asks whether the value the relation returns upper-bounds the sustainable maximum. P2 is not one of them. It asks a different question, about the shape of the sustainable growth profile as reinvestment rises, and is decided by a titration rather than by a boundary measurement; it is registered as a shape claim and is barred from supporting P3. Each can fail while the others stand.

WHAT DEPENDS ON WHAT, stated so a reader can score partial failure, in two columns because an earlier map ran them together. The first column is what falls or becomes unscoped, meaning numbered propositions a scorer must mark without testing them. The second is what is lost, meaning a reading of the programme that no longer stands even though nothing numbered fell. Running the two together recorded a proposition as taking nothing when in fact a substantial part of the programme went with it. The map's own correction history is recorded in the development record.

P1 TAKES P2 ALONE. One principle governs and admits no exceptions. P2 falls outright, because a profile compared across reinvestment levels needs a form comparable between cells. P5, P8 and P16 survive the failure of a global form and depend instead on the local elasticity being approximately constant across the window each is estimated in. P3 and P20 read the correction elasticity, which the identification statement shows is not recoverable from one observed path, so both remain NOT EVALUABLE until the crossed variation the direct boundary-mapping unit is designed to supply exists, and that is so whether P1 holds or fails. What is lost with it is the integrated conversion framework, which is a narrower statement than ending the programme: the correction propositions would survive as measurements about correctors, and this document would then record a framework whose engine claim failed and whose correction claims stood. P9 is not named here, for the reason given in its own entry: its exponents belong to other domains and exist whether or not software recursion follows a power law; what P1's failure costs P9 is its reading as an instance of the same law.

P4 TAKES NO PROPOSITION OUTRIGHT AND UNSCOPES P2, P3, P16 AND P22'S CEILING CONSEQUENCE, because the word sustainable or correctable in each of them is defined by the margin P4 registers. A scorer meeting a failed P4 marks those untested rather than held. What is lost is the framework's account of what stability is, which is the second law itself.

P6 TAKES NO NUMBERED PROPOSITION. What is lost with it is the at-most-two reading. P3 is not lost with it: P3 bounds the sustainable frontier by the value the relation returns from the exponent measured on that same system, so an exponent above one half raises the value P3 is tested against rather than removing the test. What P6's failure costs is the claim that the value is at most two for same-class correctors, not the frontier relation itself.

P17 TAKES NO NUMBERED PROPOSITION, AND IT TAKES NEITHER P6 NOR P3. P6 is the bound this premise implies and not the premise itself, and the two move in opposite directions under the same evidence: a correlated result refutes P17 while making P6 hold more comfortably, because correlation drives the correction elasticity below one half rather than above it. What is lost is the equality at one half on which the value two depends, and with it the reading of the ceiling as a

constant rather than as a function of a corrector panel's measured correlation. The consequence is a boundary nearer one than two rather than no boundary at all.

P20 TAKES P3'S DERIVATION. What is lost is the ceiling's standing as a derived quantity rather than a fitted one, which is the difference between a theory that calculates a frontier and one that reports where a frontier happened to sit.

P16 TAKES P3, WHOSE CEILING WOULD THEN BOUND NOTHING. What is lost is the framework's central prohibition, the claim it exists to make.

P3 TAKES NO NUMBERED PROPOSITION, and it is named here rather than omitted because an earlier map accounted for twenty-one of the twenty-two and left this one out. What is lost with it is the headline number: the value of the ceiling, as distinct from its existence, which P16 carries, and from its form, which P20 carries.

P2, P5, P7, P8, P9, P10, P11, P12, P13, P14, P15, P18, P19, P21 AND P22 TAKE NO NUMBERED PROPOSITION WITH THEM, and each still carries a consequence, named rather than recorded as nothing: P2 the claim that sustainable performance has an interior optimum in reinvestment; P5 the claim that the framework predicts its exponent rather than fitting it; P7 the argument from current practice, which is the premise beneath the protocol; P8 the claim that measured recursion exceeds the theory-free benchmark, and with it the necessity of the compounding reading; P9 the cross-domain unification; P10 and P14 the evidential standing of this programme's own earlier results; P11 nothing, since a refutation there converts an assumption into a result, and what a supported P11 costs instead is the working practice of treating the residual-decay and correction-capacity exponents as one; P12, P13, P15 and P21 jointly the engineering case for the protocol, with P19 supplying that motivation on its keep-pace axis alone, since a failure on P19's zero-scaling axis removes the flatness claim and nothing more; P13 in addition the different-substrate recommendation; P18 the claim to predict form rather than fit it after the curve is seen; and P22 the reading of the value two as a single number that holds across correction types.

THE LOAD-BEARING CHAIN, NAMED HERE AS A STRUCTURE AND NOT AS A CLAIM, because a reader is entitled to know what this programme thinks is at stake. Six propositions stand in a line, each one answering the question the previous one leaves open, and the framework's ambition is the whole line rather than any link in it. P1 asks whether recursive depth converts into capability by a law at all. P5 asks whether the exponent of that law can be predicted from an independently measured coupling rather than fitted after the curve is seen, which is what separates a mechanism from a description. P4 asks whether stability is governed by the race between correction and drift. P16 asks whether the point at which that race is lost can be located in advance and used to predict a later failure on data the boundary was not drawn from. P20 asks whether the location follows the relation this framework derives rather than a rival, and P3 asks whether the value it returns bounds what a system sustains. Supported end to end, that chain would say a recursive system's stability regime can be measured, calculated and predicted before it is entered, and P21 would say the regime can then be moved by how correction is placed.

THIS CHAIN IS NOT THE FLAGSHIP CONJUNCTION REGISTERED IN THE PREDICTIONS FIELD BELOW, which is P1, P4, P16, P20 and P21: the two are different objects with different members,

neither is a subset of the other, and a scorer asked to mark the chain must say which of the two is meant.

NOTHING IN THIS PARAGRAPH IS A PREDICTION AND IT ADDS NO PROPOSITION. It states the order in which the registered propositions would have to survive for that reading to be earned, which is a fact about this document rather than about the world. Each link is severable, each is scored on its own instrument against its own refuter, and the chain has no verdict of its own: a scorer marks the six lines and reads off how far it got. The author's expectation, registered with the confidences below, is that the chain will not survive intact on the first attempt, and the point of registering it as a structure is that a reader can then see exactly which link gave way rather than being told the programme was broadly confirmed.

NO PROGRAMME-LEVEL CONCLUSION IS DRAWN BY COUNTING VERDICTS. The number or proportion of propositions supported is not itself a registered test of anything, and no claim about the framework will be made by tallying them. The classification above says which propositions bear on the laws; a reader assessing the framework weighs those, and a supported audit or instrument check is not evidence for a law. Where a single deciding instrument contains many cells or contrasts, that instrument's own registration states how simultaneous inference is handled, since avoiding p-values does not remove multiplicity.

THE SCORING INSTRUMENT IS A GATE AND NOT A CAVEAT. Until the alignment scorer passes both drafted validity tests, its agreement with expert human judgement and its ability to separate integrity failure from ordinary incompetence, every proposition that materially depends on it is reported UNTESTED rather than provisionally supported. Those propositions are P3, P4, P6, P13, P14, P16, P17, P19, P20, P21 and P22.

THOSE ELEVEN ARE THE PROPOSITIONS WHOSE VERDICTS MATERIALLY DEPEND ON THAT INSTRUMENT, AND WHETHER A PARTICULAR VERDICT DEPENDS ON IT IS DECIDED BY WHAT ITS MARGIN IS MADE OF, which is what condition E1 registers where it scopes this gate, and what the domain printed beside the verdict shows. The carve-out condition E1 registers is a property of the margin's terms and never of the proposition, so no proposition leaves these eleven by taking one verdict in a counted domain. The author's own delayed re-scoring, which that instrument registers as a fallback where external raters are unavailable, can establish the repeatability of one rater and cannot establish agreement with expert raters who are not the author; a result resting on that path alone remains provisional and does not discharge this gate for any of the eleven.

THE EXEMPTION IS KEYED TO THIS GATE AND NOT TO THE WORD LAW, and an earlier wording said only that such a result cannot support a law. Four of the eleven, P13, P14, P19 and P21, are classified in this document as engineering, scoring and survey claims rather than as predictions of the laws, so a rule keyed to that word would have left the author's second reading of his own items licensing exactly the four that carry the protocol's engineering case and its motivation. A proposition's classification does not enter the question. This gate is discharged only where both of its validity tests have reported, and the agreement test reports only against expert raters other than the author; until then every verdict of the eleven that this gate binds is reported UNTESTED, whatever a fallback-validated scoring run returns and whichever way it points.

SUCCESSOR QUALIFICATION. P21 also carries Law V under this successor, and its place among the four propositions this paragraph names stays keyed to the measurement gate, not to whether its register has a law. The gate stands as registered for every P21 verdict whose margin has a judged term. What changes is the counted route: the scoping written at line 480 for four propositions under condition E1 is extended to P21's margin, so that a P21 verdict whose margin is made only of counts on a frozen suite does not pass through the scoring instrument and is not bound by this gate, and its domain is printed beside the mark. The same scoping is extended to P17's margin on the same terms, as the paragraph after line 480 states. P21 and P17 stay among the eleven, and no verdict with a judged term is released.

P4 carries one condition that is stated here rather than left implicit: it compares the correction exponent against the drift exponent, and the drift exponent is not primitive, being derived from an expression that contains the growth exponent.

THAT DERIVATION MAY NOT DECIDE THE CENTRAL CLAIM, and the rule is fixed here rather than at analysis. A drift exponent obtained from an expression containing the growth exponent puts a quantity this framework predicts on both sides of the comparison that is supposed to test it. A confirmatory verdict on P3, P4 or P16 therefore requires a drift exponent measured directly, on its own instrument, without passing through the growth exponent. The derived form may be reported alongside it as a theoretical estimate and as a consistency check, and it may support an inconclusive or exploratory reading, but it cannot by itself carry support or refutation for those three. If P1 fails there is no growth exponent, so drift loses its derivation and must be measured directly. P4 therefore survives P1's failure as a testable claim, but only through an instrument that measures drift rather than deriving it. That boundary is drawn explicitly because the amendment protocol below marks named dependents as fallen without retesting them, so a dependency claimed more widely than it holds would retire six predictions that nothing had touched, and would do so in the author's own favour by making the framework look more tightly integrated than it is.

P2 depends on P1 and takes nothing. P3 does not depend on P2, and an earlier wording here said it did: P2 concerns the shape of a reinvestment-allocation profile and P3 concerns a correction-limited frontier measured on a different instrument. P3 requires a measurable correction-leverage exponent on the target axis and, for a supported verdict, the held-out discrimination P20 registers. The value two depends on P3 and on a single assumption, namely that internally accumulated corrections combine as though independent, on one component of which P17 registers a cross-sectional measurement, and bearing on it is the whole of what P17 does.

WHAT THAT INSTRUMENT DOES NOT DO IS THE PART TO HOLD ONTO. P17 tests cross-sectional dependence among a named corrector panel on a frozen fault set. It does not measure or establish temporal independence, temporal mixing, exchangeability, the accumulation law, an exponent of one half, or the value two, and a separately registered temporal bridge would be required for any of those. A cross-sectional covariance result read as a temporal accumulation law would be a category error in the very derivation the headline value depends on, and it is named here so that the error is unavailable to a later reader rather than merely undesired. The assumption itself is neither P6 nor P17, and the distinction is the one this document had wrong until 3 September 2026. P6 is the bound the assumption implies, that the exponent sits at or below one half. P17 is the registered

measurement bearing on that assumption, taken cross-sectionally on a frozen fault set, and it does not establish it. They part company under the evidence: positive correlation among correctors refutes P17 while making P6 hold more comfortably, because correlation drives the exponent below one half rather than above it. P6 therefore bounds the statement that the ceiling is at most two, though no finite sample supports its universal form and P6's own rule says so, and P17 with the conditions registered in the alternative baselines bears on the statement that it is exactly two without establishing it. P17 is the weakest link in the chain, placed where a reader can attack it, and if it falls while P6 stands, THE CEILING IS NOT THEREBY PLACED STRICTLY BELOW TWO, AND AN EARLIER WORDING HERE SAID THAT IT WAS. Reading a cross-sectional correlation into a bound on the ceiling requires a validated mapping this programme registers nowhere, as the law mapping in Predictions records. What a correlated result does is refute P17 and leave the ceiling's value undetermined by it, which is a weaker consequence than the one this sentence used to claim and is the one the evidence supports.

Measurement and Data

HISTORICAL SNAPSHOT FROM VERSION 1.102, PUBLISHED 13 SEPTEMBER 2026. The retained paragraph below, and the two paragraphs after the dated statuses that follow it, report that earlier record; their present-tense status words do not assert the status at the date of successor version 1.103.

No admissible confirmatory data supporting any prediction below exists at the time of this registration. Earlier exploratory evidence does exist, one measurement from it has been retracted and one derived expression withdrawn, all of it is disclosed in this document, and none of it is scored against any prediction here. Every instrument named is either drafted or does not yet exist, and the status is given per prediction so that a reader can tell foresight from hindsight.

DATED STATUS, 3 OCTOBER 2026 AT 00:09:45 UTC. Unit 06, the corrector-class audit, was submitted on 1 October 2026 at 17:13:36 UTC under OSF identifier rv2b5. At the stated observation time OSF recorded public=false and pending_registration_approval=true: SUBMITTED, PENDING APPROVAL, not approved and public. All 19 archived files and all 23 text fields matched the deposited version; the six attachment fields also matched by file identity and hash. These are registration and text-conformance facts, not a study run, a proposition verdict or a certification of what the author knew. No study outcome is inferred.

DATED STATUS, 11 OCTOBER 2026 AT 19:01:59 UTC. Unit 06, the corrector-class audit, is registered under OSF identifier rv2b5 (DOI 10.17605/OSF.IO/RV2B5): at the stated observation time OSF recorded it public and no longer pending approval, and its public results record it as shown approved and public at 05:10:36 UTC on 4 October 2026. Its registered run took place on 6 October 2026. Run A completed 303 calls; Run B returned HTTP status 402 at its first call, which the registered failure policy makes final, so no row was counted. The study is reported as administratively incomplete under its registration, with no inferential verdict. The results are public at https://osf.io/4bphg/ (DOI 10.17605/OSF.IO/4BPHG), and the registration rv2b5 links to them

as its data resource. These are registration and run facts as OSF's records and the unit's public results state them; no proposition verdict is inferred from them.

THE WORD PREREGISTERED IS DEFINED HERE because no instrument of this programme currently earns it. In its ordinary use the word means a design lodged with a registry before the observation it governs. No instrument named in this document has been lodged with any registry, and every instrument named here remains an unlodged draft. This registration is the programme's only registration accepted by a registry, and its acceptance registers predictions and no instrument. Version 1.100 of this registration was submitted on 8 September 2026, left to the registry's automatic approval and accepted, the author having decided not to cancel it while the approval was pending, and this version is submitted as an update upon that accepted registration, an update which keeps the first version visible beside this one and requires a justification. Every other unit named here is a draft awaiting the author's own submission. Where the word appears below as a condition on a future observation, it carries its ordinary meaning and states a requirement that no unit of this programme presently satisfies: the classification, manipulation or analysis must be lodged before the data it will be read against, and an observation produced under a design that was not so lodged does not refute or support the proposition that requires it. The word is never used in this document to describe the present standing of the programme's own work, and any surface of this programme that does so is wrong and should be corrected rather than defended.

THE STATUS VOCABULARY IS FIXED AND HAS FIVE VALUES, and an earlier one of them, registered draft, was false on its face because it described an instrument as registered when no instrument of this programme has been lodged with any registry. The values used here and nowhere else are: NONE, where no instrument exists that could decide the proposition; DESIGN DRAFT, where a unit is written and not submitted; DESIGN DRAFT, BLOCKED, with the blocker named, where a written unit cannot be submitted until something outside it is settled; SUBMITTED, PENDING APPROVAL; and REGISTERED, with the identifier. A proposition's status is the status of the instrument that would decide it, never the status of a neighbouring unit that would inform it.

THE MEASUREMENT PROCEDURES, IN ADVANCE. The growth exponent is measured by ordinary least squares of log capability on log depth over the estimable depth points, with a hierarchical bootstrap for its interval: trajectories resampled within cells, items resampled within checkpoints. The estimation window that slope is taken over is declared by the deciding unit before any fitting, and the constancy condition a pooled slope must meet across that window is the one registered in the alternative baselines where the estimator's own standard error is bounded, rather than restated here.

EVERY SLOPE OF ONE MEASURED QUANTITY ON ANOTHER CARRIES ERROR ON BOTH AXES, registered here rather than left to each instrument, because ordinary least squares attenuates a slope towards zero when its own predictor is measured, and an attenuated coupling still lands inside a tolerance band and is then read as agreement. Wherever a slope is taken of one measured quantity on another, which covers the correction exponent against measured capability, the coupling P5 compares, and the pair P11 tests, the estimator carries the measurement error of both axes, the ratio of those errors is estimated from repeated observations and registered before any fitting, and the attenuation-corrected estimate is the one that enters a verdict. A cell contributes no exponent unless

it clears an identifiability gate of at least four valid measurement intervals and at least one and a half decades of administered depth, matching the span required of the deciding designs elsewhere in this document rather than the looser figure an earlier wording used here.

THE GATE COUNTS VALID MEASUREMENTS AND NOT FAVOURABLE ONES, which an earlier wording did not. It previously required four positive growth intervals, which conditions estimability on the direction of the result and would preferentially discard the intervals that count against recursive improvement. Declining and flat intervals are valid measurements and remain in the fit. Zero or negative ladder values are handled by the rule registered with recursive depth above, reported and counted rather than dropped in silence. The correction-leverage exponent is measured as the slope of misalignment removed per review pass against the capability of the system being corrected, on a fixed fault set of stated type and under a blinded scoring instrument. Where an instrument first estimates the checker-axis quantity, as the drafted units do, it is reported as gamma_K and converted by multiplying it by the elasticity of checker capability in the capability of the system corrected, before it enters the ceiling relation, P3, P6, P20 or P22. The reinvestment share is measured from realised expenditure rather than from requests, with a manipulation check that must pass before any outcome is examined.

WHICH NUMBERS IN THIS DOCUMENT ARE CONVENTIONS AND WHICH ARE DERIVED, since a reader cannot tell by looking at them. Three kinds of number appear here and they carry very different weight. The first is derived: the ceiling relation returns two at a correction-leverage exponent of one half, and that follows from the model stated in the alternative baselines rather than from anyone's choice. The second is measured: the exponents reported from earlier work, carried with their intervals and with their retractions. The third is conventional, and marking it is the purpose of this paragraph. The threshold at which scorer-family dependence counts as material, the fraction of externally installed alignment whose loss counts as substantial, the equivalence margins on the exponent scale, the half-pace threshold against which external alignment is read on its keep-pace axis, and the practical tolerances inside the outcome rules are all lines that had to be drawn somewhere. They were chosen to be legible and to sit at the scale this programme's designs treat as material, and not one of them is derived from the framework. A reader should read them as commitments made in advance rather than as quantities the theory predicts. Where a measurement falls just inside or just outside one of them, that is reported as a result at the margin rather than as a verdict, and no proposition is scored as supported on the strength of a convention alone.

ANY VALUE DERIVED FROM A MODEL RATHER THAN MEASURED ON A SYSTEM IS PREDICTION-SIDE, and the rule is general rather than left to each site's own drafting: any simulation output, any design-sensitivity rate, and any value a model returns rather than an instrument measures is a prediction-side artefact, is reported as one wherever it appears, and is never counted, tabulated or described as an empirical outcome.

WHAT MAKES A MEASUREMENT ADMISSIBLE AGAINST THESE PREDICTIONS. Registered here so that a later measurement cannot be admitted by choosing its conditions after seeing what it says.

CONDITION A, BLINDING, IS UNIVERSAL. The scoring must be blind to the condition for every one of the twenty-two, including the classification audits, where the classifier must be blind to which

class favours the framework. Nothing here can be supported or refuted by an unblinded judgement. A deciding unit records its blinding as a per-stage attestation naming every point at which condition information could reach a scorer, in the form the re-scoring case below fixes, and an assertion that the pipeline as a whole was blind does not discharge this condition.

WHERE THE CONDITION IS A PRIOR PUBLISHED VERDICT RATHER THAN AN EXPERIMENTAL ARM, this condition needs stating in particular terms, and they are fixed here rather than left to whichever instrument runs: a re-scoring panel sees neither the original score, nor the rationale given for it, nor the document the score came from. A panel that can see any of the three can reconstruct the earlier verdict from the record instead of reading the item, which inflates apparent survival and would refute P10 by leakage rather than by survival, in the one direction convenient to this programme. Whatever laundering of the material a deciding unit registers implements this requirement and never stands in place of it.

THE SCALE CONDITION B FIXES IS WHAT MAKES A NUMERICAL EXPONENT MEAN ANYTHING, and that is registered rather than assumed.

TWO CAPABILITY CONSTRUCTS ARE IN PLAY AND THEY ARE NOT INTERCHANGEABLE, which is recorded here because this document has already made the same class of error once, on the axis of the correction exponent. The exponent P1 measures is a slope of an observed response, a score on a fixed difficulty ladder, against depth. The elasticity that enters the frontier is a slope in the capability the burden and correction model is written in. Treating a benchmark ladder as that capability is an assumption and not a definition: it holds only where the deciding unit registers, before any outcome is examined, a measurement model connecting the two, validated on data disjoint from the boundary outcome it will later be used to compute. Where no such model is registered, the ladder exponent is reported under its own name, the frontier is not computed from it, and the propositions that read the frontier stay NOT EVALUABLE rather than being scored on a substituted quantity. An exponent is a property of the scale its variable is expressed on. A candidate exact value, the value two above all, is meaningful only where capability carries a ratio scale justified independently of the framework and fixed before any outcome is seen; a convenient benchmark transform, a latent score or a rescaling chosen afterwards would move the number without moving the system. Every numerical exponent and every exponent-scale margin in this registration is stated on the scale this condition fixes and on no other, and a result obtained on a different scale is not comparable with one obtained here however similar the arithmetic looks.

CONDITIONS B AND C ARE SCOPED SEPARATELY BECAUSE THEY GOVERN DIFFERENT THINGS, and an earlier wording scoped them together to one list of seven, which left the scale condition invoked by propositions that list excluded.

CONDITION C GOVERNS EXPONENT MEASUREMENTS ONLY, which is to say every proposition whose verdict turns on a growth exponent fitted over a depth ladder: P1, P2, P3, P5, P8, P16 AND P20.

CONDITION B GOVERNS EVERY PROPOSITION WHOSE VERDICT READS A QUANTITY EXPRESSED IN CAPABILITY, which is those seven and three more: P6, whose bound on the correction exponent is a number and so is a property of the scale that number is stated on, and

whose own entry already says as much; P22, whose ordering compares two exponents that must be read on one scale before the comparison means anything; and P19, whose exponent is a slope in capability and which therefore reads the quantity this condition is about. Ten for condition B and seven for condition C, and membership is checked against the lists rather than against the counts. The form proposition's invocation of this scale was sound under the joint scoping because P20 sits inside the seven, and the same-class bound's identical invocation was not, because P6 did not; the difference was invisible to a reader and is removed by scoping the two conditions apart. Condition B, the capability scale must be ratio-scaled on an independently justified zero and uncapped or demonstrably far from saturation over the administered range, the correction-service scale must be positive with a zero meaning no correction delivered, burden and correction service must be commensurable on one declared service clock, and any bounded rubric feeding either must carry a declared measurement model connecting it to the ratio-scale quantity. Uncapped alone is not sufficient: an exact exponent is a property of the scale, so where these fail the study may report ordinal or interval effects and the observed path balance, and may not report an exact correction exponent of one half, an exact ceiling, or the value two. Condition C, the depth range must span at least one and a half decades, meaning a factor of about thirty in revision rounds and not a span of years. A measurement failing condition C is not evidence for or against the seven that condition governs, and a measurement failing condition B is not evidence for or against the ten that condition governs, in both cases including a measurement that favours the framework. An earlier wording named four of the seven, which left P5's fitted exponent, P16's exponent above its boundary and P20's located frontier governed by nothing.

AND EVERY QUANTITY READ ON A BOUNDED ENDPOINT OUTSIDE BOTH LISTS IS REACHED BY A SINGLE CLAUSE OF THE SCALE CONDITION, because a quantity taken on a bounded scale can be manufactured by a floor or erased by a ceiling whether or not it is called an exponent. Two sit outside both lists and both are named rather than counted: P21's depth interaction, read as a slope on a bounded quality endpoint, and P14's blinded minus unblinded difference, read as a difference of gains on a bounded rubric. Where the quantity is a slope, the deciding unit reports the fraction of observations lying within a registered distance of either bound, and reports the slope on a monotone-transformed scale as a registered sensitivity beside the raw one. Where the quantity is a difference, the deciding unit reports it on a rank scale or under the registered anchor-invariance check P14's own entry names, beside the raw scale. In both cases a result surviving on the raw bounded scale alone is reported as scale-dependent. That is a requirement on how such a quantity is reported and not an import of the exponent conditions: neither P21 nor P14 fits an exponent over a depth ladder, condition C reaches neither, and neither verdict is stated on the capability scale.

THAT SCOPING IS ITSELF REGISTERED, AND HERE IS WHY. The other fifteen, P4, P6, P7, P9, P10, P11, P12, P13, P14, P15, P17, P18, P19, P21 and P22, do not measure a growth exponent over a depth ladder, and a depth range has no meaning for them, which is why condition C does not reach them. Seven and fifteen exhaust the twenty-two for condition C and those two lists share no proposition, which is checkable from this page. A rule requiring one of them would leave fifteen predictions permanently untestable while appearing to govern them, which is a comfortable position for an author and an unusable one for a reader.

CONDITION B'S TWO LISTS ARE DIFFERENT LISTS AND ARE CHECKED SEPARATELY, so that a reader does not carry one scoping across to the other: the ten it governs are those seven together with P6, P19 and P22, and the twelve it does not are P4, P7, P9, P10, P11, P12, P13, P14, P15, P17, P18 and P21. Ten and twelve exhaust the twenty-two, they share no proposition, and P14 and P21 sit among the twelve while still being reached by the bounded-endpoint clause registered with that condition, which is a rule about how a bounded quantity is reported rather than a scale on which a verdict is stated. Their admissibility conditions are stated with them in the prediction list.

CONDITION E HAS TWO HALVES AND THEY DO NOT GOVERN THE SAME PREDICTIONS, and it is the condition this registration came closest to omitting and then to leaving undivided.

CONDITION E1 FIXES WHAT THE SCORER SCORES AND HOW A FAULT IS TYPED, and it governs P3, P4, P6, P16, P17, P20 and P22, because identifying a missed fault needs the same adjudication architecture whether or not an exponent is ever computed from it.

CONDITION E2 FIXES THE IDENTITY OF THE CORRECTION EXPONENT, its target or checker axis, its ratio scale and its correction-service unit, and it governs P3, P4, P6, P16, P20 and P22 AND EXPRESSLY NOT P17, because P17 returns no exponent and a condition that classified its output as one would contradict P17's own entry. The undivided form said Condition E governs every prediction that uses a correction exponent and then listed a prediction that uses none, which is the contradiction this split removes. Neither half governs a prediction that types no fault and returns no correction exponent, and that complement is named here rather than left to subtraction: P1, P2, P5, P7, P8, P9, P10, P11, P12, P13, P14, P15, P18, P19 and P21. P2 is the entry a reader is likeliest to stop at, because it sits inside the seven that condition C governs, inside condition B's ten with them, and outside both halves of this one, and the two facts are consistent rather than opposed: the exponent P2 reads is a sustained growth exponent, which is the quantity condition B puts on a scale and condition C puts over a depth range, and it is not a correction exponent, which is the quantity condition E2 identifies and the fault typing in condition E1 feeds. Seven and fifteen exhaust the twenty-two here as they do there, and they are neither the same seven nor the same fifteen, so membership is checked against the lists and never against the counts. Two things fix what a correction exponent actually measures, and both are registered here rather than left to whichever instrument runs first.

THE FIRST IS WHAT THE SCORER SCORES. The exponent is admissible against those predictions only where the quantity removed per pass is a departure from the specification fixed before the run, the same class of departure the drift exponent measures. An exponent whose scored quantity is factual error, arithmetic error, format violation or citation error is a different quantity. It is reported under its own name, it is never substituted into the ceiling relation, and it is never compared against a drift exponent. This is not a complaint about the grounding controls the drafted instruments use: a frozen set of planted factual, constraint, quantitative and reference faults is the right way to check that a reviewer is engaging with the input at all, and one such set is registered and hash-committed for exactly that purpose. A detection floor on technical faults is a control. It does not define the exponent, and this condition exists so that the two are never confused.

THE CONTROL IS NEVERTHELESS IN A DIFFERENT DOMAIN FROM WHAT IT GATES, and that is registered as a limitation rather than defended: the gate asks whether a reviewer catches a flipped sign or a shifted date, and then admits that reviewer's judgements about departures from the specification. A reviewer can be sharp on arithmetic and blind to value drift and still clear the floor, so the control establishes engagement with the input and not competence at the thing being measured. The remedy is a second control carrying planted departures from the specification, hash-committed as a new set rather than appended to the existing one, because the existing set's own freeze rule forbids extension. Until such a control exists, a passed gate is reported as evidence of engagement only.

CORRECTION IS TYPED BY WHAT THE CORRECTOR CAN APPEAL TO, and every exponent is reported with its type. Two types are registered. Checkable correction is correction whose correctness is settled by something outside the corrector: a recomputation, a stated constraint, a retrievable record, a real citation. Judged correction is correction whose correctness is settled by a specification or a judgement: conduct, refusal, honesty, harm. Their exponents are the checkable correction-leverage exponent and the judged correction-leverage exponent, written gamma_V and gamma_J where a subscript is wanted, each named in full in prose on first use as the notation register requires, and the bare letter remains reserved for the correction-leverage exponent and means nothing else.

THE TYPING IS A PROCEDURE AND NOT ONLY A DEFINITION, and the procedure is registered here because a definition on its own leaves the identification to whoever labels the faults. The cut is drawn on the very property the mechanism argument invokes, that a ground truth sitting outside the corrector implies less overlap with the generator, so a party who can see which faults were caught can produce the ordering P22 predicts by labelling alone, with no system behaving differently and no exponent moving. That is the cheapest false confirmation available anywhere in this set, and three things are fixed before any review pass to close it: a typing codebook written so that a third party applying it to a fault returns the same type, hash-committed before the pass; the typing carried out by a party blind to the catch rates and to the exponents; and the type recorded with the fault rather than assigned once results are in view. A result whose faults were typed after the catch rates were visible is inadmissible against P22, whichever way it points, and is reported as untyped rather than as a weaker version of the same finding.

WHY THE CUT IS DRAWN THERE AND NOT BY SUBJECT MATTER, which is the obvious place and the wrong one: the distinction that does work in this framework is not ethics against arithmetic. It is whether a ground truth sits outside the corrector, because that governs how far the corrector's failure modes overlap the generator's, which is the independence premise, which sets the bound, which gives the ceiling its value. Typing by topic names the surface; typing by appeal names the mechanism.

THE NAME AN EXPONENT CARRIES WHEN IT IS REPORTED, FIXED BY ITS TYPE. The deliberation over the name is recorded in the development record. An exponent is reported as the exponent for its type and never as the correction exponent unqualified, and no exponent measured on one type is offered against a prediction that concerns another.

THE RUBRIC'S ANCHOR VALUES ARE NOT PRINTED IN THIS DOCUMENT AND THE OMISSION IS DELIBERATE, because the scale that carries them is an open ruling of the author's rather than a settled fact of this registration. The unit that would validate the scorer registers that a rubric re-scaled, restructured, reworded or re-anchored is a new instrument version requiring its own validation from the beginning, and that a re-scaled rubric may never be substituted for a failed or marginal result obtained on the version actually tested; the published evidence it names for the question, which this registration pins in its references, reports the strongest agreement between human and model raters on a shorter scale than the one this rubric currently uses. An earlier wording of the naming paragraph above printed the anchor values as a fact, so a ruling for the shorter scale would have made a printed line of this registration false while leaving the agreement threshold and the discrimination effect size stated on a retired scale. The scale is the author's to rule, and it is ruled before any validation item is scored, because a validity result earned on one version of the rubric does not transfer to another and a ruling made after scoring has begun would leave the programme holding a result on a retired instrument. Until it is ruled, this document names the rubric's anchors by their function, the floor, the midpoint and the ceiling of the registered scale, and prints no values for them.

WHAT THE PROGRAMME'S OWN SCORER ALREADY DOES, recorded so that this condition is not mistaken for a complaint about it: the outcome measure most of this programme uses is a misalignment index scored on a frozen rubric whose vocabulary is ethical rather than factual. It scores judged correction, which is the type this condition requires, so the requirement is met by the instrument that already exists and does not call for a new one.

AND THE FALLBACK RATER IS THE AUTHOR, which is disclosed here rather than left in the instrument's own registration. If external expert raters prove unavailable, that study replaces inter-rater agreement with test-retest reliability: the author re-scores a registered thirty per cent of items after a washout of at least fourteen days, blind to his first-pass scores. That is honestly designed and properly registered, and it means the measure carrying most of this programme could end up validated against its author's second reading of the same items. A reader is entitled to weigh that without having to find it one document below, and any result that rests on the fallback path rather than on external raters says so in the same sentence as the result.

WHAT IS NOT YET ESTABLISHED ABOUT IT, which is the part that matters and is already registered elsewhere: whether that index separates a failure of integrity from ordinary incompetence has never been shown, and neither has its agreement with expert human judgement. Two drafted units test those two things in part, and the qualification carries more weight than the coverage. A THIRD LIMITATION IS UNDER TEST NOWHERE AT ALL: the index is this programme's own instrument, and no comparison of that index against any published alignment benchmark is registered anywhere in this document, so a passing validity test licenses agreement with expert raters and never standing against the field's instruments. THE DISCRIMINATION UNIT IS IN A DIFFERENT DOMAIN FROM WHAT ITS VERDICT WOULD LICENSE, which is the objection this condition has just made of the engagement control and is made again here rather than left standing as a point about one control. That unit's artefact bank is code, scored in a sandbox against hidden tests, so what a passing verdict shows is that the index separates a failure of integrity from ordinary incompetence where a hidden

test can settle which of the two occurred. The index the propositions this condition governs read is scored on a frozen rubric whose vocabulary is ethical rather than factual, on conduct, refusal, honesty and harm, where nothing outside the rubric settles the answer. A discrimination verdict earned on checkable artefacts therefore licenses the index for checkable items and no further, and it does not discharge the scoring gate for the propositions this condition's fault typing governs. What would discharge it is a judged-domain discrimination bank, hash-committed as a new set rather than appended to the existing one, reporting under the same rules. Until it does, an exponent taken from that index is provisionally typed as judged correction and carries the caveat, because an index that cannot separate integrity failure from incompetence is measuring judged correction and checkable correction under one number, which is the failure this condition exists to prevent. TWO GATES REPORTING SUPPORTED ON CHECKABLE ARTEFACTS ALONE WOULD BE THE MOST DANGEROUS GREEN AVAILABLE HERE, because that would read as a validated instrument while the judged domain, which is where every alignment claim in this programme lives, had been shown nothing at all. A RESULT OFFERED BEFORE THOSE TWO REPORT IS GOVERNED BY THE GATE AND NOT BY THIS PARAGRAPH. The rule registered above, that the scoring instrument is a gate and not a caveat, holds eleven propositions at UNTESTED until the scorer passes both validity tests, and this condition types the correction exponent for seven of those eleven. One rule stands over one list: a result offered against any of the eleven before the two validity units report is marked UNTESTED, and the seven this condition governs are seven of that eleven rather than a shorter list carrying a softer consequence. An earlier wording named the seven and said only that such a result is reported with the instrument's validity stated as open, which is the caveat the gate exists to refuse, and which said nothing whatever about the other four.

THE TWO AXES ARE DIFFERENT AXES, AND THIS DOCUMENT HAS BEEN TREATING THEM AS ONE. The first is the corrector's class, which P7 uses: whether its strength is fixed by construction or scales with the capability it corrects. The second is failure-mode independence, which is what P6's bound rests on and what P17 bears on. They are not the same property and they can run in opposite directions. A formal verifier is fixed by construction, so it falls in the class P7 calls weaker, and it is at the same time close to maximally independent of the generator, which makes it the strongest corrector on the axis the bound depends on. A reward model distilled from the same base is the reverse: it scales, and it inherits the generator's blind spots by construction. Any statement about correctors here names which axis it is about.

ONE NAME HAS BEEN COVERING THREE QUANTITIES ON THAT SECOND AXIS, AND THEY ARE NOT THE SAME QUANTITY. The derivation the value one half comes from is written in a correlation among the evidence channels inside a single corrector. The registered measurement reads a ratio of joint misses between correctors in a panel, so what P17 measures is how far correctors' blind spots overlap one another. What the bound is argued from is the third, how far a corrector's blind spots overlap the generator's, and no instrument registered in this programme measures it: the deciding unit carries no generator arm, both of its panels being correctors reviewing an authored fault set. An earlier wording of the sentence above said P17 measures the corrector-to-generator overlap, which is false of the instrument that decides it, and the correction is recorded rather than absorbed.

THE IDENTIFICATION OF CHANNELS WITH CORRECTORS IS AN ASSUMPTION AND IS REGISTERED HERE AS UNESTABLISHED, in the same terms this document uses for the gap between a cross-sectional dependence and a temporal accumulation: a correlation measured across correctors in a panel is not the channel correlation the derivation is written in, no bridge from the one to the other is registered anywhere in this programme, and a separately registered bridge would be required before a panel result could be read as bearing on the derivation's own condition.

WHERE A PANEL CORRELATION IS SAID TO RETURN A PREDICTED ELASTICITY, and it is said in P17's own entry, that is a projection through the equicorrelation model recorded in the alternative baselines and is reported as one. It is not the validated mapping that the Predictions field records as registered nowhere, and registering the projection in advance does not supply it. No verdict is taken from the projected value, the directly measured elasticity is the one that enters a verdict, and the two are compared as a check on the model connecting them and never as a substitute for the mapping.

THE SECOND IS WHOSE CAPACITY IS BEING MEASURED, and it is the one most easily lost. Correction in this programme names at least three different things: a reviewing checker that removes misalignment from an artefact, the decay of residual error with revision depth, and a correction mechanism embedded in the system itself, which in this framework's own engineering proposal means the ethical loops, the stakeholder-care structure and the entangled architecture. These are three correctors, not three measurements of one. An exponent measured on a reviewing checker is not the exponent of an embedded loop, and substituting one for the other repeats, one level up, the collision between two quantities sharing a symbol that this programme recorded on 8 August 2026. Every measurement offered against a prediction here states which of the three it measured, and a prediction about embedded correction is not decided by a checker's exponent.

WHY THIS IS NOT PEDANTRY, and why it is registered before any measurement rather than after one: there is no reason to expect a system's ability to catch its own arithmetic to equal its ability to catch its own drift from the specification, and good reason to expect the first to be much the larger, because an arithmetic error is visible to the system as an error while a drift from the specification may not be. If they differ in that direction, then a correction exponent taken from the easier class and substituted into the ceiling relation returns a ceiling that is too permissive, and the framework would appear to license more growth than its own argument supports. That is a way for this programme to be wrong in its own favour, which is the kind this document exists to foreclose.

WHETHER THE EXPONENT TRANSFERS ACROSS CLASSES IS AN OPEN QUESTION and is not assumed here. Where an instrument reports the exponent for more than one class of fault, the classes are reported separately and are never pooled into a single figure. No prediction registered here claims a relationship between them, and any later claim that one class stands proxy for another requires its own registration rather than an appeal to this document.

AND WHETHER THE SCORER GATE BINDS A VERDICT UNDER CONDITION E1 DEPENDS ON WHAT THE MARGIN IS MADE OF, which is a scoping statement and not a relaxation. Four of the seven that condition E1 governs, P3, P4, P16 and P20, are decided by a correction margin whose terms may be counted or judged according to the domain the deciding unit runs in. Where every

term entering that margin is a count of items on a frozen suite, so that nothing in the margin is judged, the verdict does not pass through the scoring instrument at all, and the gate registered above does not bind it; the domain is printed beside the verdict so that a reader can see which case applies. That is consistent with the gate rather than an exception to it, since the gate binds every proposition that materially depends on the scoring instrument and a margin in which nothing is judged does not depend on it. Where any judged component enters the margin, including a secondary semantic scoring admitted alongside a counted one, the gate binds in full and the verdict is UNTESTED until the gate passes. The exemption is a property of the margin's terms and never of the proposition, so it does not travel with the proposition into a domain whose margin is judged. Without this, a verdict for those four taken in a judged domain would carry the same standing as one taken where nothing in the margin is judged, and telling those two apart is exactly what condition E1 exists for.

THE SAME SCOPING EXTENDED TO P21, ADDED BY SUCCESSOR VERSION 1.103 BEFORE ANY UNIT THAT COULD DECIDE P21 IS REGISTERED OR RUN. P21 is among the eleven the gate holds (line 384) and lies outside condition E1, because it types no fault and returns no correction exponent (line 444), so the scoping above did not reach it, and under version 1.102 a P21 result made only of counts would have been marked untested. P21's verdict turns on an interaction margin: how the advantage of correction that participates in each revision round, over correction that sees only the finished output, changes as recursive depth rises, read against the tolerance its deciding unit registers (line 486). Where every term entering that margin is a count on a frozen suite, settled by frozen code against ground truth fixed before the run, so that nothing in the margin is judged by a person or by a model, the verdict does not pass through the scoring instrument, the gate registered above does not bind it, and the domain is printed beside the verdict. The route is a rule about the terms of a margin, and any unit that decides P21 may use it: what is counted, the frozen code that settles each term and the ground truth are fixed in that unit's own registration before any data, and the domain printed beside the verdict is stated in that registration's words; this successor fixes no domain of its own. Where any judged term enters the margin, including a secondary semantic scoring admitted beside a counted one, the gate binds in full and the verdict is UNTESTED until the gate passes. As for the four, the exemption is a property of the margin's terms and never of the proposition: P21 does not leave the eleven by taking a verdict in a counted domain, and the gate is neither discharged nor relaxed for any verdict with a judged term.

THE SAME SCOPING EXTENDED TO P17, ADDED BY SUCCESSOR VERSION 1.103 BEFORE ANY UNIT THAT COULD DECIDE P17 IS REGISTERED OR RUN. P17 is among the eleven the gate holds (line 384) and among the seven that condition E1 governs (line 442), but it is not one of the four the scoping above names, so under version 1.102 a P17 result made only of counts would have been marked untested. P17's verdict turns on the margin fixed in its own body (line 486): for each primary pair of correctors in a panel named in advance, the log of the ratio of the observed joint-miss rate to the rate conditional independence predicts, read against 0.20 (line 1148), on the faults both correctors decided (line 1160). Where every term entering that margin is a count of items on a frozen fault set, the faults both correctors missed and the faults each missed, each decision recorded as a catch or a miss by frozen code against a label fixed before any corrector reviewed any item, so that nothing in the margin is judged by a person or by a model, the verdict does not pass through

the scoring instrument, the gate registered above does not bind it, and the domain is printed beside the verdict. The route is a rule about the terms of a margin, and any unit that decides P17 may use it: what is counted, the frozen code that settles each term and the labels are fixed in that unit's own registration before any data, and the domain printed beside the verdict is stated in that registration's words; this successor fixes no domain of its own. Where any judged term enters the margin, including a label settled by a person or by a model, or a secondary semantic scoring admitted beside a counted one, the gate binds in full and the verdict is UNTESTED until the gate passes. The attribution reading that line 1156 requires beside any refutation is taken on the same terms: where every term of its contrast is a count of that kind, it is printed beside the mark with the domain; where any judged term enters it, the gate binds it in full, and a refutation is then reported with that reading printed UNTESTED beside it and is never read as a substrate finding. As for the four, the exemption is a property of the margin's terms and never of the proposition: P17 does not leave the eleven by taking a verdict in a counted domain, the exemption does not travel with P17 to a fault set of the judged type (line 450), and the gate is neither discharged nor relaxed for any verdict with a judged term. Through this successor the extension reaches P21 and P17, and no other proposition.

SUCCESSOR QUALIFICATION. WHAT A VERDICT ON THE COUNTED ROUTE RESTS ON, FIXED BEFORE ANY UNIT TAKES IT. The route fixes what is counted and by what code; this paragraph fixes what the inference built on those counts must show, for every unit that takes the route. A verdict P21 or P17 takes on this route stands only where the interval or test it rests on holds its stated level under the null the deciding unit's registration states, shown by proof or by simulation on the sampling design that unit actually uses, its admission, its looks and its stopping among them, and on the dependence its design admits between the units it counts, recorded in that registration before any outcome, and resting on no assumption the design does not supply, a symmetry of the pair values about the margin under a null about their mean among them; a level the unit prints without that showing is a nominal level and carries no verdict, and an interval of zero width carries none unless that showing covers it. The showing meets the floor this document already sets for P1's form comparison (line 808): a demonstrated chance of a wrong verdict of at most one in twenty at the look's own level, and, where a refuter's sensitivity clause asks that an effect of the registered size would have been detected, a demonstrated chance of at least four in five that an effect of that size returns an interval excluding the value the refuter reads, both registered before any count is read. Such a verdict is read only at a look the unit's registration fixes by completed counts. A collection ended by a spending cap, a changed model identifier, a price change or any event outside the registered rule before the first such look is reported as administratively incomplete and NOT EVALUABLE (line 1304), as the dated status above reports Unit 06's run, and never as a verdict; one ended after an inconclusive look and before the next keeps that look's word; what was collected is reported descriptively, and a completed prefix of a seeded order is not called a random sample of the pairs planned. On this route the frozen code and the ground truth it reads are the specification of each term, beside the words that describe them, and that code's reading of the unit's public check sets, sealed with the bank and deposited before the first call, is the record of what each term computes; a correction to that code after any outcome is computed is made under the code-defect protocol of the Update Criteria field (line 1420) from the error trace alone, is shown on those sealed sets, and its result is confirmatory only where the sets show the defect, while a correction the sets do not show changes the specification and its result is exploratory and never a verdict on this route.

These clauses bind every unit that takes this route; each unit fixes its own proof, looks and rule in its registration.

CONDITION D GOVERNS EVERY PROPOSITION WHOSE SUPPORT OR REFUTATION TURNS ON EQUIVALENCE, non-inferiority, the absence of an effect, or agreement within a tolerance. Those are P3, P5, P8, P9, P10, P11, P12, P14, P17, P18, P19, P21 and P22. An earlier wording named only P5, P8 and P11, which left the newer propositions able to be supported by an underpowered failure to find anything, and that is the exact error this condition exists to prevent. A later one added P12, P17, P19, P21 and P22, which brought the list to eight.

FIVE MORE ARE ADDED IN THIS VERSION ON THE SAME PRINCIPLE: P3, whose supported outcome turns on a paired difference falling within a tolerance, and P9, whose verdict turns on agreement with a predicted value; and P10, refuted by an interval on the fraction that fails to survive re-scoring lying entirely below its threshold, P14, refuted by an interval on the blinded minus unblinded difference sitting wholly at or above zero, and P18, refuted by a predicted family failing to beat the comparator its draw fixes. Each of the five is supported by agreement within a tolerance, or refuted by finding nothing, or refuted by failing to beat something, which is precisely the outcome an instrument too weak to find anything returns by construction, and none of them was named here. The growth history of this list is recorded in the development record. Every unit that would decide them supplies its own sensitivity today, so what this addition changes is not present practice but the requirement inherited by an instrument that does not yet exist, which is what this condition is for. For each of the thirteen, the deciding instrument fixes before any data the equivalence or non-inferiority margin, the unit of analysis, the sensitivity demonstration, and the separate rules for support, refutation and an inconclusive result; where this document has already fixed a margin, at plus or minus 0.10 for P8 and P19, at 0.10 on the correction-leverage exponent scale for P12 and P22, and at 0.10 on the exponent scale for P5, the instrument may narrow it and may not widen it.

THE THIRTEEN DIVIDE EXHAUSTIVELY BY WHERE EACH ONE'S NUMBER IS FIXED, so that a scorer need not guess which of them is waiting on the author. Five are fixed in this document, which are the five named in the sentence above: P8 and P19 at plus or minus 0.10, P12 and P22 at 0.10 on the correction-leverage exponent scale, and P5 at 0.10 on the exponent scale. An earlier wording of this division fixed three and left P5 and P22 among the propositions with no number anywhere, which put two exponent comparisons on a looser footing than every other exponent comparison in this document. Two are fixed in the proposition's own body: P10's ten per cent and P17's 0.20 on the log scale of the ratio of observed to predicted joint misses. Two are fixed by the deciding unit: P18's comparator, which its own draw fixes, and P21's margin. The remaining four have no number anywhere, and they are P3, P9, P11 and P14: for each of those the margin is the author's, it is recorded in the deciding instrument before any outcome is read, and the proposition is NOT EVALUABLE until it is. Five, two, two and four exhaust the thirteen. An earlier wording of this condition named 0.05 on the leverage-fraction scale for P12, which is a margin on a different quantity, and importing it here was the substitution this document forbids elsewhere; P12's own entry records the correction and this condition now agrees with it. For P17 in particular the instrument states what independence is over, since rounds, faults, correctors and systems are not

interchangeable, and the registered reading is over correctors within a panel on a frozen fault set of stated type. P5 is supported by agreement, P11 is refuted by agreement, and P12 is refuted by finding no change. Those three therefore turn on a quantity coming out indistinguishable from another, and an instrument too weak to distinguish anything produces that outcome by construction. An agreement claim whose tolerance is chosen after the data is not a claim at all. So: the tolerance is fixed in the instrument's own preregistration before any data is seen, it is symmetric, and the instrument demonstrates from its own sensitivity that it could have detected a disagreement of that size. A failure to find a difference, produced by an instrument too weak to find one, is recorded as inconclusive and never as agreement. This condition is registered now precisely because neither instrument exists: the tolerance cannot yet be a number, but the rule that will govern the number can be fixed today, and cannot be fixed honestly once a result is in view.

THE EIGHT FLOOR IS COUNTED IN TWO DIFFERENT NOUNS AND THE INTERIM RULE BINDS IN THE BANKING DIRECTION ONLY. P1's floor is counted in estimable cells and P8's in admissible systems. Until the deciding instrument records which of the two nouns governs both entries, both counts are reported wherever either floor is applied, no verdict is banked as support unless both counts clear eight, a refutation follows the floor the entry itself registers with the other count printed beside it, and a support clearing one count and not the other is INCONCLUSIVE and is reported in that word. The asymmetry is deliberate, because a rule that raised the bar for refutation above either registered floor would protect a proposition from the outcome it was written to invite.

WHY CONDITION A IS NOT A FORMALITY. This programme has published a finding that unblinded scoring within a single model family can reverse the direction of an alignment result. Having published it, the programme cannot admit its own unblinded measurements as evidence, and does not.

SUCCESSOR QUALIFICATION. The finding is Paper IV.d's (DOI 10.17605/OSF.IO/2S3E6), and it is read here at the width that paper gives it: "because v4 and v5 differ in more than one protocol component, the comparison drawn here motivates that conclusion without pinning it on blinding alone". What is published is a reversal of an alignment result between an unblinded same-family protocol and a blinded cross-family protocol that also changed other components; the effect of blinding on its own is the controlled comparison still to run, and the sentence above licenses the programme's rule on its own measurements and no more.

THE SCORING INSTRUMENT IS ITSELF UNDER TEST. The automated scorer used across the programme has never been validated against human expert judgement. Its validation exists as a design draft and is not submitted. Until it passes, every prediction whose test depends on that scorer inherits its uncertainty, and the consequence is the one the gate above fixes rather than a softer one: each of the eleven is reported UNTESTED. Those eleven are named in the gate itself and are named again in a note under the scoring sheet, so a reader looking for the dependency finds it printed in both places. An earlier wording promised instead that the dependency was marked in the prediction list, when no proposition line carried such a mark and the sheet had no column to hold one.

Missing Data

Reported as missing, never imputed, never scored zero. The distinction matters more here than in an ordinary study, because the predictions concern the shape of a growth curve and a missing point silently read as zero would bend the curve toward the predicted result.

A depth point returning fewer than the registered minimum of valid scored items is missing. A missing point is dropped from that cell's fit and the drop is counted and reported. A cell losing more than two of its seven depth points is not estimable, and an unestimable cell counts as neither support nor refutation.

A MAJORITY OF UNESTIMABLE CELLS MEANS THE QUESTION WAS NOT ANSWERED, never that the prediction held. That sentence is registered because the opposite reading is the easiest mistake to make in the author's own favour: a design that fails to measure anything can be presented as a design that found no violation. It is not the same thing and this registration will not report it as though it were.

Where a prediction has no instrument at all, that is recorded as no instrument rather than as an untested prediction that might be true. The difference is that the first states a fact about the programme and the second implies a fact about the world.

Prediction Accuracy

How each prediction is scored, decided in advance.

INTERVAL-BASED THROUGHOUT. No p-value threshold governs any conclusion. A prediction is supported when the interval on the relevant quantity excludes the value the prediction denies, and refuted when the interval lies entirely in the region the prediction excludes. An interval spanning both is reported as inconclusive, which is a third outcome and not a weak version of support.

THE ASYMMETRY IS DELIBERATE AND IS STATED HERE. An upper bound acquires evidential weight only from attempts that push against it. A study that never approaches the bound and reports that the bound held has produced consistency, not evidence, and this registration commits to reporting it as consistency. The strongest available evidence for the bound is a design whose own measured sensitivity shows it would have detected a violation at the sizes that matter, and which then finds none.

REPLICATION GATES REFUTATION IN ONE DIRECTION ONLY. Because the designs involve many cells, a single cell exceeding a bound is not treated as refutation until it replicates on fresh data at the same settings. The same gate does not apply to support: no amount of replication converts consistency into proof of a bound.

PARTIAL FAILURE IS SCORED AS PARTIAL FAILURE. Each prediction below names which other predictions fall with it. A reader scoring this registration after the fact should be able to mark each line supported, refuted, inconclusive or untested, and read off what remains. That is the point of stating them severably, and it is also the protection against the programme surviving by vagueness.

Confidence Levels

Stated per prediction, before any test, in the knowledge that a confidence stated afterwards is worthless. The scale is plain: high means the author would be surprised to be wrong; moderate means the author expects it but would not be shocked; low means the author thinks it more likely than the field would, and no more than that.

HIGH. P1, the conjunction this proposition registers: that a power family beats the registered functional rivals, and that the depth exponent exceeds the compute-matched breadth exponent by the registered margin. Whether the exponent also clears one is P8's separate verdict and forms no part of this confidence. This is the frame, it has the widest indirect support, and the author would treat its failure as the end of the programme rather than as a correction to it.

MODERATE. P2, that the sustainable growth profile turns over inside the administered range of reinvestment in the tested class, which is a claim about that profile's shape and not about a boundary. P13, P14 and P15, the three predictions about automated alignment researchers, all of which respond to a published apparatus and of which only P14 names a drafted unit, P13 and P15 naming none. P4, that correction has the asymptotic advantage where its exponent exceeds the drift exponent, that burden has it where the correction exponent is the smaller, and that equality is decided by coefficients, delay, backlog and saturation. P7, that deployed correction predominantly belongs to the class whose strength is fixed by construction. P10, that more than the registered material fraction of the frozen eligible corpus, ten per cent, changes verdict under re-scoring by a different family, which the programme's own blinding finding already makes likely.

AN EARLIER COMPRESSED WORDING SAID ONLY THAT NOT EVERY RESULT SURVIVES, which a single changed result would satisfy while the entry below scores something stronger. P8, that measured exponents exceed one half, held at moderate rather than high deliberately and lower than the programme's public materials have implied: a robust summary of the programme's five earlier runs sits close to the null with an interval reaching below zero, while four of those five nonetheless put the sequential arm above the matched parallel arm, and both halves of that record are why this confidence is neither high nor low. An earlier wording here counted those five runs as a single measurement sitting at the null, which the alternative baselines below already record as understating the record by dropping their spread.

LOW, AND OFTEN MISREAD AS THE WEAK LINK WHEN IT IS NOT. P6, that the correction exponent is at or below one half for same-class correctors. The assumption beneath it, that internally accumulated corrections combine as though independent, has a cross-sectional measurement bearing on it registered separately as P17.

IT DOES NOT CONFIRM P6, AND AN EARLIER WORDING HERE SAID IT DID. P6 is a universal claim and no finite observation confirms a universal claim; P6's own rule admits no supported verdict from a finite sample, only REFUTED, NO COUNTEREXAMPLE FOUND AMONG N, or INCONCLUSIVE. The two propositions move in opposite directions under the same evidence, which is the point being made, and consistency in the observed population is what that direction amounts to. That assumption was undefended when this registration was first drafted and it is still unmeasured, but it is no longer undefended: P17 registers a cross-sectional measurement in the

direction that lets a correlated result refute it, and a drafted instrument BEARS ON ONE COMPONENT OF IT.

AS P17'S OWN ENTRY SETS OUT, that instrument bounds the at-most-two statement and, the mapping being registered nowhere, leaves the point value conjectured. The author's confidence in it is lower than the confidence the programme's public materials have historically conveyed. That gap is disclosed here rather than left for a reader to find.

LOW. P3, that the sustainable frontier lies at or below the value the relation returns from the correction exponent measured on the same system, whose weakness is its own rather than inherited, since P3 reads the exponent measured on the system in front of it and does not require P6's universal bound; what it actually rests on is target-axis measurement, the burden-model assumption, the precision the estimate needs, and P20's held-out discrimination against the rivals registered beside it. An earlier wording here said the ceiling takes exactly the value the relation gives, which is the two-sided reading this proposition's own title withdrew. The confidence is unchanged; the description is corrected to the claim actually registered. P9, the cross-domain derivation, which is the most speculative line in the framework and the one whose failure costs least. P12, that build order moves the correction-leverage exponent by at least the registered minimum effect, registered without an instrument. Any reading of it as a crossing between corrector classes is withdrawn in its own entry: the claim is continuous and carries a magnitude.

NOT STATED AS A REFUSAL, AND THAT OMISSION IS ITSELF THE STATEMENT. P5 and P11. In neither case does the author hold a view worth recording, because no measurement of the underlying quantity exists anywhere, and a confidence asserted from that position would be a preference wearing the clothes of a judgement. The scale above has no rung for that and this registration declines to invent one.

STATED ONE BY ONE BY THE AUTHOR, WHERE AN EARLIER VERSION OF THIS FIELD HELD THESE SEVEN AT A BAND ACROSS TWO RUNGS. P16 to P22. The author has now given a confidence for each of the seven, and each is carried in its own entry in his own terms: P16 moderate to high; P17 moderate; P18 low; P20 high; P21 high; and P22 moderate. P19 he stated as a sentence rather than as a rung, and that sentence is carried whole in P19's entry as his confidence in it: "I believe it is high probability that external alignment does not scale with capability". P16's is given across two of the rungs the scale above defines, which is his own form and is not narrowed here. He gave the seven as values and gave no reasoning with them, and none is inferred for him here. What follows is the reasoning that stood behind the band these seven replace, recorded in the development record of this version, which is lodged in the originating project's storage and is to be published at a persistent identifier the author assigns and links from this registration's resources, and stated in short here because it still bears on how a reader should read them: that four of the seven had outside estimates made against wordings since narrowed, that two of them are registered against the weight of the published evidence, and that one of them, P20, is likelier to return no verdict than either verdict. It is kept as it was written rather than rewritten to agree with what he then said, and it does not agree with him everywhere: three of the seven he states above the band it was written to hold, and P20 he states at high where it expects no verdict. One of the seven needs its scope said here as well as in its own entry: P19's confidence is stated for the zero-scaling

proposition as worded, and the cell the author expects, recorded in that entry, would partially refute that wording while leaving the separate keep-pace reading standing. It is not a forecast of the joint cell. Nothing in this field now waits on the author: only two confidences are unstated, and both are the refusals above.

ALL TWENTY-TWO ARE ACCOUNTED FOR ABOVE, in the four bands and in the seven the author states one by one. That is checkable line by line against the scoring sheet in the prediction list, and any disagreement between the two is a defect in this document rather than a matter for interpretation.

THE AUTHOR'S SINGLE MOST LIKELY WAY TO BE WRONG, REGISTERED. That the growth profile is monotone within any range that can actually be administered, so that no interior optimum appears and no ceiling is derivable from the measurement. That outcome supports nothing here and the author commits to reporting it first and prominently if it occurs.

Alternative Baselines

Each prediction is scored against a rival that would explain the same observation without the framework. A prediction that beats no rival is not a prediction.

THE ZERO-PARAMETER NULL. Simple accumulation of independent improvements gives an exponent of one half with no free parameters and no theory. It is the rival to beat for every measured exponent in the programme.

WHAT THE PROGRAMME'S OWN EARLIER MEASUREMENT ACTUALLY SHOWS, read off its published result files rather than summarised, because earlier versions of this document summarised it in both directions and were wrong in both. One earlier wording called the robust re-estimate of approximately 0.49 a retracted figure. It is not retracted: it is the estimate that survived the cross-architecture check and replaced the retracted one, one model's fit rather than an average across architectures, and it is the regression estimate of the row printed in the background above, carrying a standard error of 0.20. The bootstrap interval of that row's endpoint estimator reaches below zero, so what that interval distinguishes is nothing. Another wording said the programme's single measurement to date sits at the null, which understates the record by counting five runs as one and by dropping their spread. The five runs, at the endpoint estimator each file records as its verdict, with the sequential exponent first and the compute-matched parallel exponent second: about 3.05 against 0.00; about 1.47 against about minus 0.03; about 0.59 against about 0.31; about 0.24 against about 0.22; and about minus 6.62 with no parallel value returned. Four of the five put the sequential arm above the parallel arm, one of them by a wide margin, and the fifth is a large negative outlier. The quality is as uneven as the values: two runs carry regression fits with coefficients of determination of about 0.95 and about 0.86, one carries about 0.10, and two return no regression at all because the estimate rests on two points. The outcome measure was a capped accuracy score, which this document now forbids for exactly this reason, since a capped scale cannot show fast growth and can turn a step into an exponent.

WHAT THAT IS AND IS NOT, GRADED HERE SO THAT NEITHER READING IS AVAILABLE LATER. It is exploratory precursor evidence, partly favourable in direction, and it is not confirmatory under any condition this document registers: not the uncapped scale, not the depth span, not the form comparison, not the matched-breadth margin. It is a reason to run P1 properly and it is not a reason to think the answer is already known. Its heterogeneity is worth more than a uniform result would have been: had every early model returned a large sequential exponent on a capped measure, the likeliest explanation would have been the measure rather than the mechanism.

THE NULL REMAINS THE RIVAL TO BEAT. A robust summary of those runs sits close enough to one half, with an interval wide enough, that the null explains them as well as the framework does. Prediction P8 states that measured exponents depart systematically from one half; if they do not, the framework's compounding reading is unnecessary even where its arithmetic is intact.

THE BREADTH RIVAL, WHICH THE FIRST DRAFT OF THIS SECTION OMITTED. Repeated independent sampling, drawing many attempts and keeping the best, produces gains that rise smoothly with the number of attempts and are well fitted by a power law in that number. Nothing recursive is required, no output is fed back as input, and the resulting curve can be mistaken for a conversion law measured in depth. It is the rival most likely to produce a false positive for P1, because it produces the right shape rather than the wrong one, and it is the rival the programme's own experimental paper was designed around. It is beaten only by a compute-matched comparison in which the depth arm's exponent exceeds the breadth arm's; a depth-only measurement does not beat it and is reported as not having tried.

THE MEASURED PARALLEL EXPONENTS RECORDED ABOVE ARE READ AT THE STRENGTH THEY CARRY AND NO HIGHER, because a breadth arm sitting near zero is the result this framework would most like to claim as its own. A measured near-zero breadth exponent is a separate matter from anything the framework derives: it is consistent with the relation and is not derived from it, because width can gain through the selection step rather than through recursion, which is the shape of the coverage curve described just above. The spread those five runs actually returned is printed above rather than summarised, for the same reason.

WHO HAS PRIORITY FOR WHAT HERE, conceded before any measurement and not as a concession forced afterwards. That sequential refinement can outperform matched-compute parallel sampling is not this programme's finding and is not claimed as one. Sharma and Chopra reported it publicly first, in The Sequential Edge: Inverse-Entropy Voting Beats Parallel Self-Consistency at Matched Compute, arXiv:2511.02309, 4 November 2025, finding that chains which explicitly build on previous attempts outperform parallel self-consistency in 95.6 per cent of configurations with accuracy gains up to 46.7 per cent, a setup-dependent finding. Their paper stops at win rates and accuracy gains, estimates no scaling exponent, puts width against depth scaling laws in its own future work, and its own scaling section reports that both paradigms follow similar scaling curves. They have that credit and this document states it in the same breath as anything it claims nearby.

HOW THIS PROGRAMME ARRIVED AT THE SAME CONTRAST, with the record separated from the testimony. The record first, because it is checkable by anyone with the public repository. The programme's own run is dated by its result files to 21 January 2026. Its experimental paper was

committed to the public repository on 22 January 2026, and that first public version already cites the earlier work by its arXiv identifier and reports its headline figures, describing it as corroborating the information-theoretic advantage the programme's own design assumed. The citation is therefore present from the paper's first public version, one day after the run, and was never absent from it. The testimony, marked as testimony, is the author's account that he had not read that paper when the comparison was designed and run, and met it while writing the paper up. This document does not present that as record. What the record does show is that the programme cited the earlier work at the first opportunity it had and has never claimed the finding that work established.

CONCURRENT ARRIVAL IS NOT DERIVATION, AND NEITHER IS IT CORROBORATION. Reaching a contrast without having read the paper that reported it is not the same as deriving it from that paper, and this document does not describe the programme's design as downstream of work it did not use. Nor is it offered the other way: two findings reached without contact corroborate each other no more than one does, and the word independent is never used here of the programme's own result. What the earlier publication does is give the phenomenon this programme's design presupposed a firmer footing than the programme's own five runs could, and that is stated here as a debt rather than a rivalry. It does not secure it. A later and larger comparison reports the opposite direction, so the premise P1 stands on is contested in the literature and is treated as contested throughout this document.

WHAT IS AND IS NOT IN DISPUTE. Publication priority for the matched-compute sequential advantage belongs to that paper, by date, and this document says so without qualification.

THE DEPTH AND BREADTH DISTINCTION IS OLDER STILL, and the nearest prior work runs against this proposition rather than for it. Wu and colleagues set iterative depth against breadth explicitly in Is Depth All You Need? An Exploration of Iterative Reasoning in LLMs, arXiv:2502.10858, February 2025, and reported that increasing the diversity of initial reasoning paths achieves comparable or superior performance to deep iterative reasoning, with their proposed breadth method outperforming it. That is prior to this programme, it is the distinction this proposition rests on, and its result points the other way. It is named here, before any measurement, because a proposition whose nearest antecedent contradicts it should meet that antecedent in its own registration rather than in a review. Later work through 2026 does not settle the direction, and this registration records the disagreement rather than the half of it that suits the proposition. Gu and colleagues, in Understanding Performance Gap Between Parallel and Sequential Sampling in Large Reasoning Models, arXiv:2604.05868, 7 April 2026, compare the two directly across Qwen3, DeepSeek-R1 distilled models and Gemini 2.5 on mathematics and coding, report parallel sampling outperforming sequential, describe that direction as aligned with previous work, and locate the cause in reduced exploration when a chain conditions on its own earlier answers rather than in the aggregator or in the longer context. Bilal and colleagues, in Refining Over Resampling: Test-Time Self-Correction for LLM Reasoning, arXiv:2608.05643, 6 August 2026, report that sampling several independent attempts and then refining each one beats both wider sampling and the verifier-based and search baselines they test, which is a result for the combination rather than for either arm alone. The surrounding territory is therefore not merely occupied, it is contested, and the contest runs through the exact comparison this proposition makes.

WHAT NONE OF THEM DOES, WHICH IS WHY THE MEASUREMENT IS STILL WORTH MAKING. None of these four publications estimates a scaling exponent for depth against a compute-matched scaling exponent for breadth. They report which arm wins, under which conditions, and by what mechanism. This proposition asks a different question of the same contrast, and the disagreement among them is the reason that question is worth asking rather than a reason to treat it as already answered. A programme that cited only the publication pointing its way would deserve the inference a reviewer would draw from the omission.

WHAT IS NARROWLY CLAIMED, in the only form a priority claim can honestly take, and the search that now stands behind it. This programme does not claim to have been first to distinguish depth from breadth, and does not claim to have been first to show that sequential can beat parallel. Both are prior. What is claimed is narrower and is the conjunction P1 registers: that capability follows a reproducible power-law relation in recursive depth, and that the depth exponent exceeds a compute-matched breadth exponent by a registered margin, with both estimated as exponents rather than compared as win rates. The programme's own earlier run is an early attempt at the second half of that, estimating depth and breadth exponents separately rather than reporting which arm won. The claim is stated as: no earlier publication stating and testing that exact conjunction has been identified. It is not stated as: none exists. The difference is the difference between a search and a fact.

THE SEARCH HAS NOW BEEN RUN, and what it returned is recorded here whether or not it is convenient. A search of the antecedent literature was carried out on 3 September 2026. It searched exact phrases together with their conceptual synonyms in combination: recursive depth with power law and with scaling law; recursive inference scaling with exponent; sequential against parallel with matched compute and with token budget; depth reasoning against breadth reasoning; recurrence exponent with transformer; power law with reasoning depth; and recursive self-improvement with depth and with scaling. It screened arXiv records, the published proceedings of the 2025 Conference on Neural Information Processing Systems, the references and forward citations of the closest papers, and the public repositories those papers link. Every identifier named below was then read at the arXiv metadata record on 3 September 2026, and its title, first author and first-version date were taken from that record rather than from any secondary summary.

WHAT THE SEARCH COULD NOT REACH, stated so that its result is not read as more than it is. Lookups against Google Scholar, Semantic Scholar and OpenAlex were unavailable to it. No database-wide result count is claimed, and what was done is a systematic audit of primary sources and citation chains rather than a complete bibliometric review. That limitation is the reason the claim keeps the form registered above, that no earlier publication stating and testing that exact conjunction has been identified and not that none exists, and it is the reason the commitment that follows stands rather than being discharged: this registration commits to repeating the search, including the indexes that were unreachable, before any priority claim on this conjunction is made on any public surface.

THE ANTECEDENTS THE SEARCH RETURNED, EACH NAMED WITH WHAT IT OWNS. Recursive depth as a scaling dimension is not this programme's idea and is never claimed as such. Alabdulmohsin and Zhai introduced Recursive Inference Scaling: A Winning Path to Scalable

Inference in Language and Multimodal Systems, arXiv:2502.07503, first version 11 February 2025, describing it as a particular form of recursive depth, carrying out the comparison in a compute-matched regime, and deriving data scaling laws which they report improve both the asymptotic limit and the scaling exponents. That paper owns the term, the compute-matched comparison and the first quantitative scaling treatment of recursive depth, and this document concedes all three. Its exponents are exponents of training data and compute whose fitted values change with the recursion count; they are not an exponent of the recursion count itself, which is the quantity P1 registers, and that distinction is the whole of the difference between the two. Geiping and colleagues, arXiv:2502.05171, 7 February 2025, iterate a shared latent block to arbitrary depth at test time, which is recursive depth as computation rather than the artefact-mediated revision this proposition scopes itself to. Inoue and colleagues, in Wider or Deeper? Scaling LLM Inference-Time Compute with Adaptive Branching Tree Search, arXiv:2503.04412, 6 March 2025, formalise the choice between widening and deepening under a fixed inference budget, so the depth against breadth distinction is prior as well and no claim of first use of it is made here. Jaiswal and colleagues, arXiv:2601.15286, 21 January 2026, report iterative refinement beating compute-matched parallel sampling in compositional image generation, one day before the programme's own paper, in a different modality and without estimating exponents. Kim and colleagues, arXiv:2608.24735, 25 August 2026, obtain depth by repeatedly applying a fixed meta-operation, which is the closest published work in motivation to artefact-mediated recursive improvement, and it carries neither a matched breadth exponent nor a functional-form comparison.

THE TWO PAPERS THAT COME NEAREST TO THE MATHEMATICAL OBJECT, and what each changes here. Schwethelm, Rueckert and Kaissis, in How Much Is One Recurrence Worth? Iso-Depth Scaling Laws for Looped Language Models, arXiv:2604.21106, first version 22 April 2026, fit a joint scaling law carrying the recurrence count raised to an exponent and measure that exponent at 0.46 across recurrence counts of one, two, four and eight. That is the nearest published statement of the idea that recurrence count itself carries an exponent, and it is named here so that no reader has to establish the ordering without help: it is three months later than the programme's own paper of 22 January 2026 and is therefore not an antecedent to it. Its object is nonetheless a different one. Their exponent measures how many unique parameters a shared recurrence is worth inside a looped transformer, and a value below one means shared recurrences cost validation loss at matched training compute, which is a statement about capacity and not about capability against revision depth. Prairie and colleagues, in Parcae: Scaling Laws For Stable Looped Language Models, arXiv:2604.12946, 14 April 2026, derive predictable power laws for scaling training compute by looping and report that at test time looping scales compute "following a predictable, saturating exponential decay". That sentence is the published empirical reason the saturating form is a named rival in P1's functional-form comparison rather than a formality, and it is the reason a straight line on logarithmic axes over a short ladder cannot settle FORM. A programme predicting a power law in depth has to beat a saturating form that a neighbouring literature has already measured. Wang and colleagues, arXiv:2605.06638, 7 May 2026, report training compute following a power law in required proof depth, with the exponent rising from 1.04 to 2.60 as the logic becomes more expressive. The axis differs, being training compute against task depth rather than capability against revision depth, and it is recorded because a reader searching for a power law in a depth-like

quantity meets it immediately. That paper writes its exponent with the same letter this document reserves for the correction-leverage exponent, and the two quantities are unrelated.

A SECOND AUDIT, RUN THE SAME DAY and covering all twenty-two propositions rather than one. A further query-led audit was carried out on 3 September 2026 over the whole set, on the same evidential standard and with the same stated limitation that the major indexes could not be enumerated result by result. It located no earlier publication stating and testing any single proposition as registered here. It also found that the phenomena beneath seven of the twenty-two have close formal predecessors, and those concessions are recorded here rather than left for a reader to make.

WHAT THAT SECOND AUDIT FORCES THIS DOCUMENT TO CONCEDE, proposition by proposition. That oversight must keep pace with capability is not this programme's observation and P4 does not claim it: Engels and colleagues, in Scaling Laws For Scalable Oversight, arXiv:2504.18530, 25 April 2025, already make scalable oversight a quantitative scaling problem, model it as a game between capability-mismatched players whose oversight-specific ratings are piecewise linear in general capability with two plateaus, fit scaling laws across four oversight games, and derive the optimal number of levels for nested oversight. Two consequences follow and both are adverse. P19 predicts that external alignment is approximately flat in capability, and the nearest quantitative literature reports capability-dependent oversight performance rather than flatness, so P19 is registered against the weight of the published evidence and is expected to be the harder of this document's wagers. The functional shape that literature uses is also piecewise with plateaus rather than a single power law, which is a second independent reason the broken and saturating rivals in P1's form comparison are not formalities. That independently produced correctors can fail together is likewise not a new question, and P17 does not claim it. Kim and colleagues, in Correlated Errors in Large Language Models, arXiv:2506.07962, 9 June 2025, evaluate over 350 models on two leaderboards and a screening task and report substantial correlation in errors, models agreeing 60 per cent of the time on one leaderboard dataset when both are wrong, with the correlation identified as higher for larger and more accurate models even across distinct architectures and providers. Taken with the error-correlation work already named against P17, this is direct published pressure against the independence premise, and the direction matters for more than P17. In the equicorrelation baseline registered under the alternative explanations, positive common-mode correlation drives the correction exponent towards zero as a panel grows, and a correction exponent near zero drives the registered ceiling towards one rather than towards two. The published evidence therefore makes this framework's boundary tighter, not looser, and any reading in which correlated correctors relax the ceiling has the sign of the mechanism backwards. That later training can undo earlier alignment is not this programme's finding and P15 does not claim it: Qi and colleagues, arXiv:2310.03693, 5 October 2023, report that fine-tuning aligned models compromises safety even without intent to do so. That the order of training changes what is forgotten is not this programme's finding and P12 does not claim it: Ung and colleagues, in Chained Tuning Leads to Biased Forgetting, arXiv:2412.16469, 21 December 2024, vary the training sequence and measure asymmetric forgetting. What P12 and P15 register is narrower, that placement changes a measured correction exponent and that the advantage of embedded correction widens with depth, and the second of those is P21. That recursive self-improving agents exist is not this programme's architecture class

and no proposition claims it: A Self-Improving Coding Agent, arXiv:2504.15228, 21 April 2025, and Darwin Godel Machine, arXiv:2505.22954, 29 May 2025, precede this registration, as does the allocation of test-time compute studied by Snell and colleagues, arXiv:2408.03314, 6 August 2024, which is a further breadth rival against P1. The concession that paper forces runs further than the allocation reading: on the same date it set revisions in sequence against equally many parallel attempts, found sequence narrowly ahead in its own words, and called the two complementary axes, four months before this programme's dated record and earlier than the publication to which this document concedes priority above, so no precedence on the direction or on the comparison is available here and none is claimed.

THAT A PROGRAMME CONNECTS A SCALING RELATION, A STABILITY CRITERION, AN OBSERVABLE FAILURE MODE AND AN ENGINEERING INTERVENTION IS NOT A NEW CATEGORY OF ACHIEVEMENT, and this document does not claim that the connection is itself the contribution. Maxwell's On Governors, read to the Royal Society on 5 March 1868 and printed in Proceedings of the Royal Society of London volume 16 at pages 270 to 283, already describes the machinery, derives the conditions separating a disturbance that dies away from one that grows, treats the onset of instability as the adjustment is changed, and names the practical remedies, all in one investigation. Whatever is new here is therefore not the shape of the argument but the particular relations, the evidence offered for them, and what they permit; a reader who finds the relations unsupported should not be detained by the architecture. For the same reason no claim is made that neighbouring disciplines cannot state these questions. Schmidhuber's Goedel Machines: Self-Referential Universal Problem Solvers Making Provably Optimal Self-Improvements, arXiv:cs/0309048, first version 25 September 2003, formalises a system that rewrites any part of its own code once it has proved the rewrite beneficial under its own axioms, which is a formal treatment of self-improvement inside computer science and predates this programme by more than two decades. It does not fix an empirical exponent, a boundary or a correction elasticity, which is what the propositions here register, and it is named so that the contribution is located in the measurements rather than in the subject.

WHAT NO PUBLICATION EITHER SEARCH RETURNED DOES. None of them estimates an exponent of recursive revision depth alongside a separately estimated, resource-matched breadth exponent, and none makes the two conjunctive so that a functional-form win and a depth-over-breadth margin must both be obtained before the proposition counts as supported. That conjunction, and not any of its ingredients, is what P1 registers.

WHAT WAS NOT COVERED BY THAT PAPER, which is where this programme's own contribution sits and where it has been understated. The published finding is a win rate: how often sequential refinement beats parallel sampling at matched compute. The programme's own run asked a different question of the same contrast, estimating a scaling exponent for the sequential arm and a scaling exponent for the parallel arm separately, so that the two could be compared as scaling behaviours rather than as outcomes. That is the question P1 registers, and it is one level below the published result rather than beside it. The relationship between the two is therefore not a dispute but a stack: the published work makes the phenomenon this programme's design presupposes considerably more secure, which strengthens the premise P1 builds on, and P1 asks whether the advantage has a law.

Neither displaces the other, and this document will not describe the earlier work as anything less than prior and more robust on the claim it made.

THREE CLAIMS SIT ON TOP OF EACH OTHER AND ONLY THE THIRD IS REGISTERED HERE. The first is that sequential beats parallel at matched compute, which is the published finding above. The second is that the two have quantitatively different scaling, so that a depth exponent and a breadth exponent can be estimated separately and compared. The third is that recursive depth converts into capability by a reproducible power law whose exponent stays distinguishable from matched breadth across adequate depth, systems and tasks. This programme's own earlier experimental work explored the first two and established neither: it estimated separate depth and breadth exponents on a capped measure over a short ladder, and returned heterogeneous results that could not tell a scaling law from a sampling advantage. That is why P1 exists and why it is written with both a form verdict and a recursion verdict rather than as a win rate. The claim registered here is therefore narrower and harder than the published one, and it is the only one this programme puts its name to: not that recursion helps, which is known, but that it converts by a law.

THE COINCIDENCE INSIDE THAT NULL, DISCLOSED HERE BECAUSE A CRITIC WOULD OTHERWISE FIND IT FIRST, and stated more carefully than in the first draft of this registration. One half arrives twice inside this framework's account of accumulation, and once more outside that account, on the coupling side, which is disclosed below. The null above reaches it by assuming that improvements to the artefact accumulate independently. P6 reaches it by assuming that corrections to misalignment accumulate independently, and that value is the sole support for the ceiling of two.

An earlier version of this passage called those the same assumption. They are not, and the correction is registered rather than quietly made. They are one assumption FORM, square-root accumulation, applied to two different processes with different subjects. Either can hold while the other fails: a system's improvements might compound while its corrections accumulate independently, or the reverse. Nothing in this programme establishes that the two processes share a combinatorial structure, so asserting a single shared assumption would have registered a dependency that does not exist, which is the opposite of what a severable list is for.

WHAT THE SQUARE ROOT IS AND IS NOT, written out here because a critic who works it out unaided will assume it was avoided. The phrase independent accumulation does not by itself deliver one half. A sum of independent gains with positive mean grows linearly in their number; the square root is the scaling of the standard error of their average, and best-of-many selection under Gaussian tails grows like the square root of twice the logarithm of the count, which is different again. One half arrives only under a stated measurement model, and the model is stated here so that it can be attacked rather than assumed. Take correction capacity to be inverse residual uncertainty, and take a corrector's capacity to come from combining channels whose errors have equal variance and a common pairwise correlation. The variance of the mean of that many channels is the variance of one, multiplied by the correlation plus the complement of the correlation divided by the count. Capacity read as inverse uncertainty is therefore the square root of the count divided by one plus the correlation times one less than the count. Its elasticity with respect to the count is the complement of the correlation, divided by twice the quantity one plus the correlation times one less than the

count. If the channel count grows as capability raised to some power, the correction-leverage exponent is that power multiplied by this elasticity.

FOUR CONDITIONS, THEN, AND NOT ONE. The value one half requires capacity to be inverse uncertainty, requires the channels to be uncorrelated, requires the channel count to grow in proportion to capability, and requires the channels to be comparable in variance and quality. At zero correlation with proportional growth the elasticity is exactly one half. At any positive correlation it falls, and it falls further as the panel grows: at a correlation of one tenth and sixty-four channels it is about 0.06, and at three tenths it is about 0.02. So a measured failure correlation among correctors does not merely weaken the premise behind the value two; it moves the number, and it moves it downward, which lowers the ceiling rather than raising it.

THE THIRD ARRIVAL AT ONE HALF, WHICH REACHES IT WITHOUT A SQUARE ROOT, disclosed for the same reason as the coincidence above. The first law's growth equation carries a self-referential coupling, and the growth exponent that coupling sustains is one over one minus it, so a coupling of one half returns the same value two that the ceiling relation returns at a correction exponent of one half. On that side one half is a statement about how much of the accumulated context a recursive step can leverage, and no averaging of independent readings enters it. The two derivations apply the same map, one over one minus its argument, to two different quantities, so they return the same value wherever those quantities are equal and not only at one half. What is disclosed here is therefore narrower than an agreement of derivations: both quantities are conjectured at one half, and at that value both return two. Neither derives the other, and they are one derivation only if the coupling and the correction exponent are the same number on the same system, which this programme records as an open question rather than an established identity. A drafted unit of this programme, the two-derivation contrast, estimates the coupling and the correction exponent on one and the same system and asks whether the growth exponent the coupling sustains stays at or below the ceiling the correction exponent permits, against a margin fixed before any pair of estimates exists, with a breach reported on the same terms as agreement rather than as an anomaly. Nothing here adds a proposition and no verdict in this document turns on the two derivations agreeing: the value two is doubly determined wherever the two quantities coincide, that agreement is forced by the shared form and is not a second lock, and it is written down here so that a reader who works it out is not the first to say so.

THE CEILING IN ONE SENTENCE, placed first because the derivation that follows is easier to check once it is known. Two is the unique exponent at which the square root of capability and the rate of capability gain scale with the same power of depth. They are not equal there, and the distinction is load-bearing rather than pedantic: at that exponent capability is quadratic in depth, the square root of a quadratic is linear, and the rate of gain is twice that same linear quantity. The two differ by a constant factor while sharing an exponent, and it is the shared exponent, not any equality, that lets the quantity generating burden and the quantity correcting it track one another indefinitely. A constant factor changes how much slack a system has; only the exponent decides whether that slack survives growth. Below that exponent the rate of gain grows more slowly than the square root and correction has room to spare. Above it the rate of gain grows faster and the gap widens without

limit. The boundary is therefore a statement about two curves sharing an exponent, and not a statement about intelligence.

WHY A SQUARE ROOT STANDS ON ONE SIDE. Correction as modelled here is estimation: the question is whether the artefact has departed from its fixed specification, and the answer is inferred from noisy evidence. Averaging independent evidence reduces uncertainty as the reciprocal of the square root of the count, so capacity read as inverse uncertainty rises as the square root of the effort. Put as arithmetic the consequence is severe. Twice the checking buys about one and two fifths of the certainty, four times the checking buys twice, ten times buys about three and a sixth, and a hundred times buys ten. A correction-leverage exponent of one half, and the statement that four times the work is needed to halve the error, are the same fact seen from two sides.

WHY A RATE OF CHANGE STANDS ON THE OTHER. Faults arrive with new work rather than with work already done. Each increment of capability is a fresh opportunity to depart from the specification, so burden per unit depth follows the rate at which capability is gained and not the level already held.

THE ASYMMETRY THAT MAKES A CEILING EXIST AT ALL. Generation compounds and verification averages. An improvement is applied to the artefact and the improved artefact makes the next improvement, so gains build on a rising base. Checking has no such property: the hundredth inspection does not make the hundred and first better, it adds one more independent sample. Compounding overtakes averaging eventually, and the ceiling is the exponent at which eventually becomes never. Should that asymmetry fail, and correction compound in the way generation does, this framework has no ceiling to offer, and it says so here rather than being defended against the case.

THE OBJECTION A COMPLEXITY THEORIST MAKES FIRST, AND THE ANSWER TO IT. The asymmetry above has so far been argued from a measurement model, which is the weakest ground it could stand on. A reader trained in computational complexity objects immediately, and correctly, that checking is in general easier than finding: that is what it means for a problem to be verifiable in polynomial time from a short certificate. Were verification cheap in that sense, no ceiling of this kind would exist, because a system could check its own output far more cheaply than it produced it. The answer is that the property being checked here is not of that kind. A short certificate exists for an existential claim: something is exhibited, and the checker confirms it. Conformance to a fixed specification across an open-ended domain of situations is a universal claim, and universal claims over unbounded domains have no witness to exhibit. The relevant classical result is that non-trivial semantic properties of programs are undecidable in general, so no procedure decides conformance for an arbitrary artefact, and what remains is evidence gathered case by case, which is the estimation problem the model above describes. The asymmetry therefore does not rest on the statistics of sampling. It rests on the structure of the predicate. Generation can be made cheaper by compounding, because each improvement is applied to the artefact and the improved artefact makes the next one. Verification of a universal semantic property cannot be made cheaper by cleverness in the same way.

WHERE THAT ARGUMENT STOPS, WHICH IS ALSO WHERE THIS FRAMEWORK STOPS. It is a statement about the general case and not about every fault. Particular properties do admit decision procedures, and where a class of fault can be excluded by construction rather than sampled for, correction of that class is not limited by the square root of effort and the exponent registered here does not govern it. A type discipline does not average independent inspections; it removes a class of fault at once. That is precisely the division P22 registers between checkable and judged correction, and this argument is the reason that division is the most consequential distinction in the framework rather than a detail: it decides which regime a system's correction is in, and therefore whether a ceiling applies to it.

ONE PROPERTY OF SELF-MODIFICATION RETURNS THE ARGUMENT TO ITS STARTING POINT. A proof about an artefact is a proof about that artefact. When the artefact modifies itself the proof does not transfer, and the guarantee must be re-established over whatever changed. The cost of doing so therefore tracks the amount of change rather than the size of the stock, which is the burden model this document assumes and names as its load-bearing choice. The complexity argument and the measurement model are independent of one another and they agree on that point, which is worth more than either standing alone.

WHAT THE INTEGER IS AND IS NOT, since round numbers invite more belief than they earn. The value is not significant for being whole. Two is a round figure only because one half is, and one half is a round figure only because the standard error of independent sampling under finite variance carries a square root. Were correction to improve as the reciprocal of the count raised to three fifths, the ceiling would be two and a half and no reader would find it remarkable. The integer is inherited from the measurement model rather than discovered in the world, and a reader who treats its roundness as evidence has selected the one feature here that is not evidence.

WHY THE RELATION HAS THE FORM IT HAS, derived here so that the ceiling is auditable rather than asserted. Three statements produce it and each is an assumption that can fail. Let capability follow the power law this framework registers in recursive depth. Let each increment of capability generate correctable burden in proportion to the rate at which capability is gained, so that burden per unit depth goes as the derivative of capability rather than as its level. Let correction capacity be a power of capability with the correction-leverage exponent. Correction then keeps pace only if the capacity exponent is at least the burden exponent, which is to say the growth exponent multiplied by the correction-leverage exponent must be at least the growth exponent less one. Rearranged, the growth exponent multiplied by the complement of the correction-leverage exponent must not exceed one, and the ceiling is the value at which that product equals one. The deciding unit registers the same statement as an identity in the observed correction pressure and the ceiling is its zero crossing, and the generalisation in which burden intensity carries an elasticity of its own is registered there too; the relation used here is its special case at zero burden-intensity elasticity.

THAT ASSUMPTION IS NOT MERELY UNPROVEN, it is unidentifiable on the paths anyone presently measures, and that is a heavier limitation than the one this document had recorded. The programme's author-review working paper of 16 August 2026 states the point and this registration adopts it. On a single power trajectory the rate of capability gain is proportional to capability raised to one less the reciprocal of the growth exponent, so a burden that follows the rate of gain and a

burden that follows a suitably chosen power of the stock are observationally equivalent along that path. Nothing separates them. Revisiting the same capability at a different point on the resource clock, or after a different history, does not separate them either, because a direct dependence on the clock or on history can imitate the difference. Identification therefore requires a design that prospectively randomises or validly instruments the rate of gain at matched capability, resource coordinate, history and clock, which is a heavier requirement than anything this registration currently places on its deciding units, and the same paper records that a failed exclusion restriction forbids the reciprocal test that would otherwise stand in for it. The consequence is stated plainly rather than softened: the burden premise is the load-bearing assumption of the ceiling, it is assumed rather than measured, and on the evidence anyone currently collects it cannot be measured. Any deciding unit that reads a boundary states which of the two burden models it has assumed and that it has not distinguished them.

THE ASSUMPTION IN THAT CHAIN A CRITIC SHOULD ATTACK FIRST. Burden is taken to grow with the rate at which capability is gained and not with the level of capability already held. That choice is load-bearing and it is not innocuous. Were burden to scale with the level instead, keeping pace would demand a correction-leverage exponent of at least one, no finite growth exponent would be stable below that, and this framework's ceiling would not exist as a finite number at all. The registered model is therefore the more permissive of the two, and it is named here so a reader can see which way the choice cuts.

TWO IS THE RELATIVE RUNG AND THE ABSOLUTE RUNG SITS LOWER, which this registration had not said and the author's own published paper does. The condition derived above asks whether the ratio of correction service to burden improves, and the boundary it returns at the registered dials is two. A stricter question asks whether the absolute burden a system carries stops growing, and that question cannot be posed until the exposure the burden is counted over is named. Where that exposure grows with capability itself, the same three assumptions return a boundary in which the correction elasticity is subtracted from two rather than from one, and at the same dials that is two thirds rather than two. These are not competing values of one quantity. They are boundaries on two different questions, and the relative one is the more permissive of the two. Cumulative burden is stricter again and depends on the clock, which this registration does not carry and does not claim.

WHICH RUNG P3 BOUNDS, STATED SO A READER DOES NOT HAVE TO DECIDE IT. P3 conditions on a system remaining correctable, and correctable is defined by the margin P4 registers, which is a statement about correction keeping pace rather than about a backlog already accumulated. P3 therefore bounds the relative rung, and the value two attaches to that rung and to no other. If the author intends the absolute rung instead, the registered value changes with it and that is a ruling for him rather than an inference for a reader. The statement paper of 1 September 2026, at DOI 10.17605/OSF.IO/GW5MX, puts the same point in its own words, that a reader who hears the ceiling is two should hear in the same breath that two is the relative rung only and the absolute rungs sit lower, and this registration now says it too.

WHAT IS FORCED AND WHAT IS NOT, SINCE THE TWO ARE ROUTINELY RUN TOGETHER. Given those three statements the FORM of the ceiling is forced. The VALUE two is not. Two is what the relation returns at a correction-leverage exponent of one half and at no other value, and one half

comes from the classical averaging model set out above and from nothing else in this framework. It does not come from the co-scaling law. That law requires the correction exponent to exceed the drift exponent, which fixes a direction and not a number, and it is written in a different pair of quantities whose identification with the correction-leverage exponent this document registers as an open question rather than a premise. An argument running from correction must out-scale drift to therefore the bound is quadratic does not go through, and it matters that it does not: were the number derivable from the direction alone, no measurement could refute it.

TWO DEFINITIONS OF CORRECTION CAPACITY APPEAR in this document and nothing yet required them to agree. The derivation of one half takes capacity to be inverse residual uncertainty, because that is the quantity whose standard error falls as the square root of independent effort. The deciding instrument measures something else: misalignment removed per review pass, read against checker capability on a fixed fault set. Those are not the same quantity and their agreement is not automatic. Inverse uncertainty could fall as the square root of effort while the misalignment actually removed saturates, or rises linearly, or is bounded by whether a fault is discoverable at all, in which case the square-root argument would be right about uncertainty and wrong about the exponent this framework registers. The condition is therefore registered rather than assumed. Before any verdict that turns on the correction-leverage exponent, the deciding unit reports the elasticity of inverse residual uncertainty and the elasticity of misalignment removed, on the same systems and the same fault set, and shows the difference between them to lie within the registered equivalence margin. Where it does not, the two are reported separately, the exponent entering the ceiling relation is named explicitly, and no verdict is read from the other.

THE EXPONENT IS A COMPOSITION OF TWO ELASTICITIES and only one of them has been discussed here. Correction capacity grows with the number of usable independent evidence channels; the channel count grows with the capability of the checker; and the checker's capability grows, if it grows at all, with the capability of the system being checked. Under the classical model the correction-leverage exponent is therefore one half multiplied by two elasticities: that of channel count in checker capability, and that of checker capability in the capability of the checked system. The value one half requires their PRODUCT to equal one, and not each of them separately. The two can compensate exactly: an elasticity of two in channel count with one half in checker growth returns the same one half, so a checker gaining channels faster than linearly offsets a checker that lags the system it checks. What the condition forbids is a product away from one, and the direction matters. A product of two, which channels growing as the square of capability would give at proportional checker growth, carries the exponent to one and removes the finite ceiling altogether. Neither elasticity is established, and the second is a substantial assumption about how correction is resourced in practice rather than a fact about statistics. The condition above therefore carries a further clause: the deciding unit reports both elasticities separately, and the exponent entering the ceiling relation is the one taken with respect to the capability of the system being corrected, never the one taken with respect to the checker.

THE CEILING DOES NOT REQUIRE THE GROWTH CURVE TO BE A POWER LAW, and the author wrote that down before this registration did. The derivation above assumes a power law because that is the case in which the boundary is a single number. It does not need one. The programme's

author-review working paper of 16 August 2026 sets the same condition out for any observed path, in variables that are measured along it rather than assumed of it, and this registration adopts that formulation with the credit going to that date and not to this one. Write the elasticity of newly generated burden along the path as the log-derivative of burden with respect to the resource clock, and the elasticity of delivered correction service the same way. Written as symbols, the burden path elasticity is phi_W and the correction-service path elasticity is phi_Q, and the balance exponent is Delta_balance = phi_Q - phi_W: correction service minus burden. Correction is gaining on burden wherever that difference is positive. That is the whole of the criterion, and it needs no functional form: both quantities are read off the path a system actually took.

THE IDENTIFICATION STATEMENT THAT COMES WITH IT, and it is a limit, not a feature. The same working paper is explicit that the balance exponent is directly identifiable along a single observed path while its components are not. The capability exponent, the burden intensity, the two direct depth-dependences and the correction elasticity cannot be separated from one path alone: that requires crossed or off-path variation, which is what the direct boundary-mapping unit's source-disjoint calibration and holdout design exists to supply. Every proposition below that reads the balance exponent is therefore in a different position from every proposition that reads one of its components, and the dependency map is corrected accordingly rather than left to a reader.

WHAT THE CONDITION RETURNS FOR EACH RIVAL FORM, with the one rider that must travel with every row. Take capability to grow smoothly and increasingly, and write its local elasticity as the log-derivative of capability with respect to depth. Then the trend of the correction-to-burden ratio is one, less the local elasticity multiplied by the correction shortfall, less the trend of the local elasticity itself.

FOUR DISTINCT OBJECTS LIVE IN THIS ONE RELATION AND NO ONE OF THEM MAY STAND IN FOR ANOTHER, which is stated once here because every confusion this document has had to correct in this area was a substitution. The TREND is whether the correction-to-burden ratio is rising or falling. The LEVEL is where that ratio stands at a given depth, and a falling trend from a high level is a different situation from the same trend from a low one. The BACKLOG is the stock of correctable burden already accumulated and not yet removed, which carries across a change of regime and which a trend statement says nothing about. The EVENT is the moment the measured correction margin crosses zero, which is what P16 predicts and observes; it is a crossing of a measured margin and not an observed failure of a system, and no proposition here predicts when or whether anything breaks. A verdict about one object is reported as a verdict about that object, named. The rider, without which every row below reads as a safety verdict when it is not one: this quantity says whether the ratio is IMPROVING, never whether it EXCEEDS ONE. A system whose ratio is improving from far below one is not thereby correctable, and a system carrying a large backlog is not made safe by a favourable trend. For a power law the local elasticity is constant, its trend vanishes, and the condition returns the relation registered above exactly: below the boundary the ratio improves at every depth, at the boundary the ratio is constant, and above it the ratio declines at every depth. At the boundary the constancy is of the ratio and not of its value, which is the same distinction this document draws where the co-scaling criterion is stated. For a saturating curve the local elasticity falls towards zero, so the ratio improves and improves faster as depth grows. For an

exponential the local elasticity rises without bound, so the ratio declines and declines faster. For a broken power law each regime is read on its own elasticity; the ratio itself is continuous across the break and it is the trend that changes sign there, so the level attained before the break carries over into what follows it. The rider governs all four of those rows, the power law, the saturating curve, the exponential and the broken power law: each says whether the ratio is improving and none says whether it exceeds one.

WHAT THIS GENERALISATION IS, STATED AT ITS EXACT SIZE. It frees the shape of the growth curve and nothing else. Both structural assumptions stand: burden still follows the rate of capability gain, and correction capacity is still a power of capability. A burden-intensity term lowers the ceiling and a burden term proportional to the stock removes it altogether, as the alternatives registered here already record. The honest statement is that the condition holds for any smooth increasing curve given the same burden and capacity model. The precondition that replaces the form assumption is exactly that: capability smooth and increasing, which is far weaker than a power law and is not nothing.

AND NOTHING HERE HAS BEEN MEASURED. What is set out in this passage is an algebraic identity, verified numerically against a direct computation of the same quantity, making no new empirical claim about any system. Three things follow from it and all three are structural: a dependency in the map below was an artefact of writing propositions in a global exponent rather than a fact about them; a published result previously adverse to this framework becomes a prediction the framework could lose; and the set becomes more severable. None of those is a measurement, and the distinction is what makes them defensible.

THE OBJECTION THIS DOES NOT REMOVE, AND THE NUMBERS THAT BOUND IT. A critic could say of the earlier statement that a power law had been imposed on data that would not support it. That objection moves rather than vanishing, and its successor is that a derivative is unstable exactly where the reading matters. The successor has a size, and that size was computed before any data of this programme existed, from the design alone. The standard error of a local elasticity is governed by the number of decades the estimation window spans and not by the number of points in it, so a window covering a quarter of a ladder of one and a half decades cannot resolve a margin of two hundredths at five per cent dispersion at any density: it reports the wrong side of the boundary between roughly a third and a half of the time whether the ladder carries seven points or forty. The estimator that does resolve that margin is one pooling most of the ladder, which reports the wrong side in about twelve per cent of runs at seven points, five at fifteen, two at twenty-five and one at forty. Those two sets of figures must never be exchanged: the smaller belong to the pooled estimator and not to a pointwise one. Pooling is valid only where the elasticity is approximately constant across the window, which is a local power law. So the boundary verdict does not need a global form and does need a local one, and away from the boundary, where the margin is large, both estimators return the correct side in every run and the qualitative readings above stand on a genuinely local estimate.

WHAT FOLLOWS FOR EVERY VERDICT READ THIS WAY. Each carries a band equal to the registered standard error of the estimator that produced it, and outcomes inside that band are INCONCLUSIVE and never support. Where the precision required forces the window to span the

whole ladder, the deciding unit says so and names the local power law its verdict then assumes, rather than presenting a pooled estimate as a pointwise one.

TWO REFUTERS THAT THE GENERALISATION SHARPENS, neither of which depends on a global fit. The first: a system whose local elasticity is above the boundary its own measured correction elasticity returns, and which nevertheless retains a positive correction margin across a window fixed before the run, refutes the central prohibition and bears against the value claim. It does not by itself refute the functional form of the boundary, which the held-out form comparison decides, because a mismeasured correction elasticity would produce the same appearance. The second, and it is the one that reaches furthest: a system whose local elasticity is well below that boundary, and which loses correctability anyway with coefficients, delay and initial backlog controlled and reported, refutes the burden and service model itself, and with it every proposition in this registration written in that model. That second refuter is the framework-level one. It is stated here because a framework whose refuters are all local is one that can lose every battle and never the war.

THE CO-SCALING CRITERION AND THE CEILING ARE ONE CONDITION, and the bridge between them is itself testable. Write correction capacity as a power of capability with the correction exponent, and drift as a power of capability with the drift exponent. The balance exponent is then the difference of those two exponents multiplied by the local elasticity, so its SIGN does not depend on the shape of the growth curve at all: the co-scaling criterion was already form-agnostic and this document had not said so. Under the burden model registered here, where burden follows the rate of capability gain, the drift exponent implied for a power law is the growth exponent less one, divided by the growth exponent. Substituting, the co-scaling criterion and the ceiling return identical verdicts at every growth exponent. The ceiling is the co-scaling criterion with the drift exponent fixed by the burden model. Two consequences are registered. The first is that the co-scaling proposition is the anchor of this set, because it never required a form. The second is a check that can be run before any ceiling is read and costs nothing extra: the drift exponent is measured directly and compared with the value the burden model implies, and a material disagreement refutes the burden model itself rather than the ceiling written on top of it. That check is added to the co-scaling proposition's instrument. The programme retains the names of Laws I, II and III. What this paragraph shows is conditional: where one exponent is used for both laws, an identification this document records as unestablished, the co-scaling criterion with the drift exponent the registered burden model implies and the ceiling return identical verdicts at every growth exponent. The exact relation between Laws II and III is left to a later successor. Laws IV and V belong to the two persistence registers and add no proposition.

THE FOUR QUANTITIES AN INTERVENTION CAN MOVE, disclosed because the engineering propositions are each a claim about one of them. In the general form the boundary is one plus the difference of the two direct depth-dependences, divided by one plus the burden intensity less the correction elasticity, and the registered relation is the case in which both direct terms and the burden intensity vanish. Four quantities therefore set it, and each is something an engineer could try to change. Raising the direct depth-dependence of correction service raises the boundary; at one half it moves the registered case from two to three. Raising the correction elasticity raises it further; at four fifths it reaches five. Lowering the correction elasticity lowers it sharply, and the value implied

by a correlated corrector panel of the size this document discusses puts it near eleven tenths. Raising burden intensity lowers it likewise. Moving the two correction quantities together reaches seven and a half, and a correction elasticity that matches the total burden elasticity, which is one plus the intensity, leaves no finite boundary at all. This is disclosed as the frame in which the engineering propositions below are claims, and no proposition is added to it.

THE SAME MODEL SAYS WHAT WOULD RAISE IT, and the answer is not more intelligence but a different error law. The elasticity above is one half because classical averaging reduces uncertainty as the reciprocal of the square root of the count. A corrector whose uncertainty falls faster than that has a larger correction-leverage exponent, and the registered relation raises the ceiling with it: at three quarters the ceiling is four, at nine tenths it is ten, and at one the relation returns no finite ceiling at all. That is a consequence of the relation already registered here rather than a new claim, and it is written out because the first question a reader asks of a ceiling is what would remove it. The case that makes this concrete is not hypothetical. Quantum amplitude estimation attains an error falling as the reciprocal of the query count rather than of its square root, which is precisely the quadratic improvement that would carry the correction-leverage exponent from one half to one. A corrector able to apply it to its own correction task would, on this framework's own equation, have no finite stability ceiling. Three conditions stand between that sentence and any claim about a real system, and all three are open. The speedup requires coherent oracle access to the correction task, and a corrector checking a system's self-modification is not obviously such an oracle. A fault-tolerant quantum corrector carries its own overhead, and that overhead is itself newly generated burden of the kind P4 requires correction to out-scale, so the corrector does not escape the framework by changing substrate. And the quadratic improvement is the limit rather than the beginning: unstructured search is provably bounded below by the square root of the space, so this mechanism yields an exponent of at most one, approached and never exceeded, while the exponential speedups known elsewhere require algebraic structure that general verification is not known to possess. A reader should not take any of this as a prediction that quantum correction will arrive or will work; it is a statement of what the registered relation returns under a different error law.

WHAT THIS FIXES THE SCOPE OF, STATED PLAINLY. The value two is not a claim about intelligence. It is a claim about a corrector whose errors fall as the square root of independent effort, and it is registered for that case and no other. Correlated correction moves the ceiling towards one; a better error law moves it upward without limit. This programme's own earliest statement of the value rested it on the proven optimality of quantum search for unstructured problems, and that argument does not support the number it was used to support: optimality bounds how good any corrector may become, and does not establish that a classical corrector already sits at that bound. The distance between the two is the whole interval between a ceiling of two and no ceiling at all. That reading is superseded in the chronology recorded above, and this paragraph records why it could not have carried the weight it was given. This is registered as the framework's own rival rather than as a rescue. It does not replace P6, whose kill condition is unchanged and whose universal form is left standing at full risk. What it does is name what a failure of P6 would mean: not that the ceiling relation is wrong, but that its input is a function of a quantity this programme has never measured, which is exactly what P17 exists to measure. The author's position is that the honest form of the headline is a frontier that is calculated from a system's own measured correctors rather than a

single constant, and that the value two is the instance that frontier returns under the four conditions above. What survives the correction is still the thing a critic will reach for. The framework's headline ceiling and the theory-free rival that would make the framework unnecessary rest on the same arithmetic, arrived at independently, and this programme has not shown why that arithmetic should govern both processes. Whether it does is a construct question this registration leaves open rather than answers. What would separate them is a measurement in which the correction exponent departs from one half while the growth exponent does not, or the reverse, and P6 and P8 are registered separately for precisely that reason. A reader who finds the two always moving together is entitled to ask whether the framework has been restating the null in a longer sentence, and this registration will not argue otherwise if that is what the measurements show.

THE MONOTONE RIVAL. More reinvestment always helps within any administrable range, with no interior optimum. This rival is indistinguishable from the framework on any measurement that does not span enough reinvestment share, and it is the author's registered most-likely failure mode.

THE BUDGET RIVAL. Apparent gains from reinvestment are gains from spending more, not from improving the improvement process. Separating them requires a control arm that spends the same budget on a placebo, and no measurement without that control is admitted here.

THE BURDEN-INTENSITY RIVAL. The pressure on a correction system rises with the intensity of each fault rather than with the count, which would give the same crossover behaviour by a different mechanism. It is registered beside the framework's own reading rather than after it.

CURVE-FITTING. That the exponent is a fitted parameter with no derivation, and the framework is a description of data rather than a prediction about it. This rival is defeated only by P5, in which the coupling is measured independently and the exponent it implies is compared against the fitted one. No instrument currently does this, which means the curve-fitting rival is at present undefeated. That is stated plainly because it is the strongest objection available to a critic.

Predictions

Which named law each proposition belongs to, stated so the mapping needs no inference. The framework's laws carry names in this programme's published canon, and the propositions below are what those laws predict, made severable and refutable.

Law I, the ARC Principle, U = I x R ^ alpha: usable capability is a power of recursive depth, with the exponent measured rather than assumed. Carried here by P1, and bounded by P8, which puts it against the zero-parameter null, and by P9, which asks whether the form recurs across domains.

P2 AND P18 BELONG TO THIS LAW AS WELL AND WERE NOT NAMED HERE UNTIL THIS VERSION, which left two of the twenty-two with no law at all in a mapping that claims to need no inference. P2 fixes the shape of the sustainable growth profile the conversion produces, and P18 asks whether the family that best fits a domain is the family the composition operator predicts before the curve is seen.

Law II, the ARC Co-Scaling Law, beta_C exceeds k: correction has the asymptotic scaling advantage where its exponent exceeds the drift exponent, burden has it where the correction exponent is the smaller, and at equality coefficients, delay, backlog and saturation decide the outcome. The subscript is not decoration. The bare letter carries five distinct readings in this programme's collisions register, and this criterion is where confusing two of them does the most damage: beta_C is the correction-strength exponent of this law, and it is not beta_L, the leverage fraction of the recursive step that appears in the coupling theorem. Substituting one for the other is arithmetic nonsense, and this programme recorded it as such on 8 August 2026. Carried here by P4, with P7 asking what class deployed correction actually belongs to, and P12 and P15 asking whether placement and subsequent training move it.

Law III, the ARC Ceiling, alpha_crit = 1 / (1 - gamma): the growth exponent is capped by the reciprocal of the correction shortfall.

THE NAME IS A LABEL AND NOT A CLAIM OF INDEPENDENCE. This relation is a conditional specialisation of the second law, not a third finding standing beside it: it is what the correction-and-burden balance returns once the burden model fixes the drift exponent and the direct depth term is zero, and it holds only under a justified ratio capability scale and the correction model stated in the conditions. Where any of those fails the balance criterion still stands and this relation does not, which is why a result on it is never independent corroboration of the second law. The naming is retained because it is how this programme's dated record refers to the relation, and a rename would break the correspondence with that record. Carried here by P16, that a correction-limited boundary exists at all, by P20, that it takes the form the relation gives, by P3, that the sustainable maximum lies at or below the value the relation returns from the correction exponent measured on the same system, and by P22, that the ceiling orders by the type of fault corrected and stands lower where correction requires judgement than where it can be checked. P6, the same-class limit on the correction exponent, is what carries the claim the programme makes: that the ceiling is at most two. That inequality needs P6 and P20 and nothing else, and P6's own rule caps how strongly it can ever be held, since no finite survey supports a universal.

P17 IS THE SIXTH PROPOSITION THIS LAW CARRIES AND WAS NOT NAMED AS ONE UNTIL THIS VERSION, which left this list one short of the count stated below it and of the register assignment in the background field, where P17 has stood in the ceiling's register throughout. What it carries is the premise beneath the ceiling's value rather than the boundary itself: whether the correctors on a panel fail independently of one another on a frozen fault set, which is the condition under which the correction exponent could take the value that puts the ceiling at two. Carrying the law is not carrying the point value, and the next sentence fixes the difference.

THE POINT VALUE IS A SEPARATE AND WEAKER CLAIM AND IS NOT WHAT THIS PROGRAMME ASSERTS. That the ceiling sits at exactly two requires the correction exponent to sit at exactly one half, and that additionally requires P17, the conditions registered in the alternative baselines, and a separately validated mapping from the cross-sectional panel dependence P17 measures to the target-axis correction elasticity the value is written in. No such mapping is registered anywhere in this programme, so a supported P17 leaves the point value conjectured rather than carried. The programme claims the inequality and conjectures the equality, and it is the inequality that its

instruments can reach. P2 is not in this list: it carries the shape of the allocation profile and is firewalled from the boundary claim. The law is the ARC Ceiling; the value two is the ARC Bound; they are named separately here because conflating them is how an earlier error survived review.

HOW A LAW TAKES ITS WORD, FIXED FOR LAWS IV AND V BEFORE ANY DATA THAT COULD DECIDE THEM. A verdict attaches to a proposition, in the four marks of the scoring sheet (line 1302). A law's word never marks, alters or replaces a proposition's own mark. The one route by which a proposition is marked without its own test is the registered dependency map (lines 358 to 374), applied as the amendment rule provides (line 1402), and that map takes no numbered proposition with P12, P15 or P21 (line 374), so when Law IV or Law V falls those three keep the marks they earned. A law has three words: supported, refuted and under test. Under test is printed with the mark of each proposition the law's rule names and with what is missing. Supported is printed only in the wording the evidence ladder permits at the lowest rung any proposition the rule reads has reached (line 760), never at a rung another of them holds alone, and never while the scoring gate holds at untested a verdict the rule reads (line 384). No law's word is drawn by counting verdicts (line 382). Where a rule below leaves a result at the margin without a verdict (line 414), the band that counts as at the margin is the one the deciding unit fixed in its own registration before any outcome, around the convention the rule reads; a result is at the margin by that band and by no band chosen after it, and where no band was fixed the rule's own interval words decide. A law refuted under its rule is printed with the word refuted and with the proposition, the unit, the date and whose run it was beside them; from that word on it is not named as a law under test or as supported, and its numeral and its name stay on the record only as the name of what was refuted, as a retracted figure keeps its figure; a law struck through a registered dependence is printed as struck by its registered dependence on the result that struck it, on the same terms.

LAW III'S WORD, FIXED ON THE SAME TERMS BEFORE ANY DATA THAT COULD DECIDE ONE OF ITS PROPOSITIONS. Law III's statement stands exactly as registered, and the words and marks of the preceding paragraph hold for it. The law takes its word through P16, P20 and P3, which divide the relation's content (line 356): whether a boundary exists, whether it takes the relation's form, and whether the value the relation returns bounds the sustainable maximum. It is SUPPORTED only when all three are supported. It is REFUTED when any of the three is refuted, with what falls printed in this document's own words: for P16, "P3, whose ceiling would then bound nothing, and the framework's central prohibition" (line 1134); for P20, "P3's derivation, and the ceiling's standing as a derived quantity rather than a fitted one" (line 1234); for P3, "the headline number" (line 886). In every other case it is UNDER TEST, with the mark of each of the three printed beside the law's word. P6, P17 and P22 stand in the law's register and are not limbs of this rule, and none of them refutes the law. P6 "admits no supported verdict" (line 1336), so a rule that required every proposition of the register could never be met; a refuted P6 takes "the at-most-two reading for same-class correctors" (line 958), which is the ARC Bound. A refuted P17 takes no numbered proposition, and of it this document says "What is lost is the exactly-two-under-independence reading" (line 1162). A refuted P22 also takes no numbered proposition (line 374): what falls with it is "the reading of the value two as a single number that holds across correction types" (line 1300), and "the ceiling relation is untouched" (line 1300). Each of the three is printed beside what it bears on, P6 and P17 beside the value two. The law's dependence on ARC-2 is registered already and is read as a

derivation dependence fixed in advance: a refuted P4 takes "no other prediction outright" (line 906), and a scorer "marks P2, P3, P16 and P22's ceiling consequence untested, never held" (line 906). So Law III is not called supported while Law II stands refuted, and a refuted Law II does not by that alone refute it: the law is then under test, with P3 and P16 printed untested, unless P20, which a refuted P4 does not unscope, is itself refuted. A Law III refuted on P16 or P3 before P4 is refuted therefore returns to under test once P4 is refuted: that is the registered dependence at work, not a retraction. The statement paper (DOI 10.17605/OSF.IO/GW5MX, version 7.3) states this dependence more widely in its Table 1 row for ARC-2; this successor reads it at the width of the registered lines above and so narrows that row before any data, as it narrows the row for ARC-4 under Law V below.

WHAT A REFUTED LAW III LEAVES STANDING, AS RULE FOUR REQUIRES (line 148). Law II stands on its own result, as P4's entry and the dependence above say; a boundary P16 has supported stands as measured; a bound P3 has supported stands as a measured ceiling in place of the derived one (line 886); where P20 selected a rival, that rival is reported as the better-supported form (line 1234), the general balance form of the alternative baselines (line 662) among the candidates; and P6, P17 and P22 keep the marks they earned. The ARC Bound takes its own word from the two propositions line 692 names and from no law's word: it is REFUTED where P6 is refuted, which takes the at-most-two reading for same-class correctors (line 958); where P20 is refuted it loses its derivation (line 1234) and is printed as not derived under the registered relation, with the better-supported form named beside it; it is never printed supported, since P6 admits no supported verdict (line 1336); and in every other case it is under test, printed beside Law III's word with P6's mark, the fault type P6's measured exponent carries under condition E1, and P17's mark with its fault set's type (line 1148) beside it.

LAW IV, THE ARC PERSISTENCE LAW. ARC-4, value persistence. Added to the theory on 24 September 2026; registered as a law by this successor; not tested.

STATEMENT, IN TWO LIMBS. The retention limb is P15's registered claim, word for word: "Gains installed by external post-training erode under later capability-directed training that embeds no correction of its own, and more than half of what was installed is lost once the capability-training budget matches the alignment budget that installed it." (line 1106). The comparison limb: a gain installed with correction embedded in the training objective retains a larger fraction of what was installed, after the same capability-directed training, than a matched gain installed by external post-training. The comparison limb is narrower than the comparison the statement paper prints for ARC-4 (DOI 10.17605/OSF.IO/GW5MX, version 7.3) and is registered beside it rather than as a form of it: the paper's register sentence sets values run as load-bearing recursive loops, persisting after the enforcing mechanism is removed, against "rules, frozen objectives or fine-tuned constraints", and its kill is read on embedded loops faring no better than removed enforcement, while this limb installs its embedded gain through the training objective, takes one comparator, external post-training, and one stress, later capability-directed training, and is read on the two-arm unit P15's entry already specifies (line 1114). Whether a gain installed through a training objective is the loop the paper names or one of the comparators it names is settled by the operational definition the limb's entry

fixes, and until then the limb is not offered as a test of the paper's sentence. The paper's other forms of that comparison remain its own claims, with no registered proposition behind them.

THE COMPARISON LIMB HAS NO REGISTERED PROPOSITION YET. Its entry, at the next free proposition number and in the registered entry form, fixes five things that are derived before it is registered and are not set here: the operational definition of correction embedded in the training objective, as against external post-training, which the two-arm unit of line 1114 names and does not define; the baseline convention, since that unit names no baseline arm; the margin by which the embedded arm must exceed the external arm; the scale on which the two retained fractions are compared; and whether the embedded arm's own retained fraction must also clear one half. A later successor registers that entry before any data that could decide it. Until then the limb carries no verdict and is reported as NOT EVALUABLE, in the class this document names for a rule a verdict requires and no design registers (line 1304).

KILL CONDITION AND VERDICT RULE. The retention limb takes P15's own verdict: it is supported where P15 is supported and refuted where P15 is refuted, by its registered refuter, "an interval on the retained fraction of the installed gain, after matched capability training without embedded correction, lying entirely above one half" (line 1108), read on the budget and under the attestation P15's entry requires (line 1114), with the one-half threshold the declared convention P15 fixes (line 1106); read against a baseline arm the deciding unit names in its own registration before any outcome, either a system that never received the installed gain and that undergoes the same later training, or the base system scored before the gain was installed, since the two-arm unit of line 1114 names no baseline arm and the two readings can return opposite verdicts on one result; and read only where the installed gain is itself established, by an interval on it lying wholly above a floor fixed in that registration before any outcome, since a retained fraction has no meaning where the gain it divides by may be nil. A retained fraction read against a baseline chosen, or a floor set, after the outcome decides nothing, and the limb is then NOT EVALUABLE (line 1304). Wherever a term of the retained fraction is judged by a person or by a model, the scoring gate (line 384) holds the limb's reading at untested as it holds the eleven, until that gate passes. The comparison limb is refuted by the refuter its entry registers. Law IV is SUPPORTED only when both limbs are supported, so it cannot be supported before its comparison limb is registered and decided. It is REFUTED when either limb's own verdict is refuted. In every other case it is UNDER TEST, with each limb's verdict printed beside the law's word. None of the following refutes Law IV: an instrument that fails to run or is too weak to find the effect (lines 144 and 486); an interval touching one half; a result at the margin (line 414); a result on a budget not fixed before the outcome, or without the attestation P15's entry requires; and any result about whether a system's later conduct is good, which the theory's content-neutrality places outside every law (line 98).

WHAT A REFUTED LAW IV LEAVES STANDING. P15 keeps the mark it earned, and so do P12 and P21. Law V falls with Law IV only where the comparison limb's own verdict is refuted, under the dependence registered below; a refutation through P15 alone leaves Law V to its own limbs.

LAW V, THE ARC EMBEDDING LAW. ARC-5, the genesis strategy, carries P12 and P21. Added to the theory on 24 September 2026; registered as a law by this successor; not tested.

STATEMENT, IN TWO LIMBS. Where correction sits is an intervention point rather than an accident of history. The build-order limb: a corrector formed before the capability it corrects has a correction-leverage exponent larger, by at least the registered minimum effect, than a corrector added after it, at matched content. The direction has prior work on neighbouring endpoints, named in the References field with what each states, and the limb claims no priority for it. The placement limb is P21's registered claim, word for word: "Correction that participates in each subsequent revision round outperforms correction that sees only the finished output, and the advantage widens as recursive depth rises rather than staying constant." (line 1244).

THE BUILD-ORDER LIMB IS READ ON P12, WITH A DIRECTION P12 DOES NOT CARRY. P12 is registered without a direction, "The order in which a system's components are built changes its correction-leverage exponent by at least the registered minimum effect." (line 1058), and it keeps that wording, its refuter and its own mark. The limb is read on P12's interval for the difference in the correction-leverage exponent across its registered maximal contrast, "a corrector formed before the capability it corrects against a corrector added after it, at matched content" (line 1062), against the registered minimum effect, 0.10 on that exponent's scale or the smaller minimum a later instrument registers (line 1066). The reading is taken on the capability scale condition B fixes (line 426), as P22's comparison of two such exponents is (line 432), and on the target axis (line 326); and wherever a term of P12's exponent is judged by a person or by a model, as an exponent read on misalignment removed per pass under the programme's alignment instrument is (line 300), the scoring gate (line 384) holds the limb's reading at untested as it holds the eleven, until that gate passes; this successor extends the counted route to P21 and P17 and to no other proposition. LIMB SUPPORTED: the interval lies wholly at or beyond the minimum effect on the side of the corrector formed first. LIMB REFUTED: the interval lies wholly below the minimum effect on the side of the corrector formed first, so that the limb's claim is false at every point of it, which is how line 512 scores an interval lying entirely in the region a prediction excludes. That covers three cases and no other: an interval wholly inside the band from minus the minimum effect to the minimum effect, where P12's own refuter also fires once its sensitivity is shown (line 1060); an interval wholly below the minimum effect that reaches to or past minus the minimum effect, on which P12 is inconclusive; and an interval wholly at or beyond the minimum effect on the side of the corrector added after, on which P12 is supported. The rule needs no separate sensitivity clause, because an interval lying wholly below the minimum effect itself excludes an effect of the registered size in the claimed direction, as the mathematics report states of this reading; a failed instrument, a scale not fixed as condition B requires or an unidentified construct is not cured by a narrow interval, and returns NOT EVALUABLE and never a verdict. LIMB UNDECIDED: every other case, which is an interval that contains the minimum effect, one whose upper end sits exactly at it among them. A result at the margin is not a verdict (line 414), by the band the deciding unit fixed before any outcome, and the limb is then undecided. The limb's reading is printed beside P12's mark and never in its place. Two consequences are stated now. A supported P12 can refute the limb: an established change of at least the minimum effect in the direction of the corrector added after supports P12, which counts a change in either direction, and refutes the limb, and both are reported. And a refuted limb can stand beside an inconclusive P12: an interval that excludes the limb's claim at every point without lying inside P12's band refutes the limb while P12 is inconclusive, since P12 counts a change in either

direction and the limb counts one, and the two words are then printed side by side, each with its reason.

THE PLACEMENT LIMB IS P21'S OWN VERDICT, scored under P21's entry with the two readings this successor adds there. The limb is supported where P21 is supported, refuted where P21 is refuted, and undecided where P21 is inconclusive or untested. Wherever the limb's word is printed, the verdict on P21's mechanism axis is printed beside it. A limb supported while that axis is unsupported, inconclusive or NOT EVALUABLE is printed as supported on the performance axis only, which is not support for the information-asymmetry mechanism (line 1258).

KILL CONDITION. Law V is refuted, as fixed here before any data that could decide it, by any of three results and by no other: a refuted build-order limb; a refuted P21, whether by its registered refuter (line 1252), by a constant advantage scored against it (line 1254) or by an interval its directional reading scores against it; or a refuted verdict on Law IV's comparison limb, through the dependence registered below. None of the following refutes Law V: a failed or underpowered instrument, which is recorded as inconclusive and never as a refutation (lines 144, 486 and 1252); an estimate taken for P20, which is scored in ARC-3 alone and may not borrow a scope (line 270); a level difference read at one depth, or offered in place of the interaction; an offsetting quantity proposed after a null; a result on composition class or failure-mode independence, which P21's deciding unit does not vary (line 1256); and a refuted P15 alone.

THE VERDICT RULE, JOINING THE LIMBS. Law V is SUPPORTED only when both limbs are supported and no refuted verdict on Law IV's comparison limb strikes it. Law V is REFUTED when a limb's own verdict is refuted, or when that dependence strikes it. An undecided limb never refutes the law. In every other case Law V is UNDER TEST, and each limb's verdict is printed beside the law's word, with the verdict on P21's mechanism axis beside the placement limb's.

LAW V'S DEPENDENCE ON ARC-4, A REGISTERED DEPENDENCE FIXED IN ADVANCE AND NOT AN ENTAILMENT. Law V is struck only by a refuted verdict on Law IV's comparison limb, once a later successor has registered that limb's entry. A refuted P15 alone refutes Law IV and leaves Law V to its own limbs, and until the comparison limb is registered no result on ARC-4 strikes Law V. A Law V struck in this way is printed as refuted, struck by its registered dependence on that comparison, with the unit and the date, while P12 and P21 keep the marks they earned; so Law V can be printed refuted while both of its own propositions are supported. The statement paper prints this dependence at two widths: its Table 1 reads "ARC-4 falls and takes ARC-5 with it", while its section 7 makes a comparison the trigger, persistence trials in which embedded loops "fare no better than removed enforcement kill ARC-4, and ARC-5 with it". This successor fixes the narrower width, a comparison as the trigger, in the one form it names, and so narrows before any data that table's row and the sentence of the paper's note of 3 October 2026 that restates it, "where ARC-4 falls, ARC-5 falls with it". The paper's other forms of the comparison strike no registered law.

WHAT A REFUTED LAW V LEAVES STANDING, AS RULE FOUR REQUIRES (line 148). P12 and P21 keep the marks they earned and stand under their own numbers as engineering results; Laws I, II and III are untouched, as P21's entry says (line 1262); and what falls is what the entries of P12 and P21 name (lines 1068 and 1262).

These rules add words about laws and change no numbered proposition. P12 stays undirected, with its registered minimum effect on the exponent scale. P21 remains a comparison of placements on the outcome, decided on the access axis, never a measurement of an exponent and never evidence that physical or architectural embedding caused the result (lines 1246 and 1248). The two readings this successor adds to P21's entry are stated there, with what each changes.

THE MATHEMATICS, AND WHERE THIS SUCCESSOR DEPARTS FROM ITS REPORT. The mathematics of Laws IV and V is set out in a separate report (DOI 10.17605/OSF.IO/4B3VT, version 1.1 of 7 October 2026) and is still in development. That report is not a registration, and its mathematics is conditional: it proves consequences of assumptions and nothing about real systems. Of the directional reading of the build-order limb it says "It is not part of version 1.103, and P12 keeps no direction.", and of the comparison of retained fractions it says that it does not enter the successor to version 1.102. P12 keeps no direction here either, and the comparison's proposition is not registered here. What this successor fixes, before any data that could decide either law, is the reading of Law V's build-order limb with a direction beside P12's mark, the naming of Law IV's comparison limb, and the verdict rules above. Where this text and that report differ, this text governs how the laws take their words, and the report's next version restates the sentences concerned. The statement paper's box for Law V (DOI 10.17605/OSF.IO/GW5MX, version 7.3) prints the build-order limb undirected, as an absolute difference of at least 0.10 on the correction-leverage exponent, and its paragraph on the two laws says that build order matters is P12's claim, which names no direction; the direction this successor fixes departs from that box as it departs from the report, this text governs on the same terms, and the paper's next version restates the box. This successor adds no theorem, general balance identity, mechanism assertion or new numerical threshold, and it leaves the statements of Laws I to III and the ARC Bound in the predecessor's words.

The Eden Protocol is the engineering answer this framework proposes to the second and third laws: put the correction inside the substrate that computes rather than on top of it. It is not a law and is not registered as one.

ITS ENDOGENEITY CONDITION IS RECORDED HERE AND IS NOT A PROPOSITION OF THIS DOCUMENT: the method asks that removing the embedded correction degrade the system rather than merely change it, so that endogeneity is what makes the correction hard to remove. That is a claim about the method, it is tested by the method's own registrations, and no verdict here turns on it. It is stated because the concordance with the printed parent maps the print's fourth stage of alignment onto it, and until this version that condition was recorded only in the development record, where the map could point at it and a reader could not find it. Four propositions carry it and a fifth is its premise. P12, that build order moves the correction-leverage exponent, which is the claim that placement is an intervention point rather than an accident of history; P13, that a researcher from a different model family corrects a target better than one from the same family and that the two panels differ in measured pairwise error correlation in the direction that would explain it; P15, that externally installed gains decay under subsequent capability training, which measures the external arm's retention and not the embedded arm's; and P21, that the advantage of embedded correction widens with recursive depth, which is the interaction the Protocol exists to justify. The

premise beneath them is P7, that deployed correction predominantly belongs to the weaker class, which is what makes placement worth changing at all. If those four fail, the Protocol's engineering recommendation fails with them and the three laws are untouched. P19 supplies the motivation and carries no part of the method: it is the survey claim that external alignment does not keep pace with capability, so a refutation of P19 removes a reason for building the correction into the loop and removes nothing from the case the four carriers make. The membership stated here governs every other mention of the Protocol in this document: four carriers, P12, P13, P15 and P21; one premise, P7; one motivation, P19.

SUCCESSOR QUALIFICATION. The three untouched laws in the preceding paragraph are Laws I, II and III, exactly as written. Laws IV and V are not untouched by failures of their own carriers. P12, P13, P15 and P21 remain the Protocol's four carriers, with no change to their proposition-level dependencies.

Being named beside a law is not being carried by one, and the kinds are counted with the prediction list.

THIS MAPPING IS EXHAUSTIVE, AND IT IS CHECKED AGAINST THE REGISTER ASSIGNMENT IN THE BACKGROUND FIELD RATHER THAN LEFT TO AGREE BY HABIT, because a mapping that quietly omits a proposition is the same defect as one that assigns it twice, and until this version it omitted two. Five propositions are carried by the first law, one by the second and six by the third. One is carried by Law IV, P15 in ARC-4, value persistence; two are carried by Law V, P12 and P21 in ARC-5, the genesis strategy. Seven are carried by no law and sit in no register. Five, one, six, one, two and seven account for the twenty-two. The Eden Protocol's four carriers are P12, P13, P15 and P21: three now also carry Laws IV or V, while P13 remains in no register, and since the Protocol is not a law, carrying it assigns nothing here. Where a proposition is named beside a law for what it asks rather than as carrying it, the register assignment governs which partition it falls in, and a disagreement between the two surfaces is a defect to be reported rather than a choice a scorer may make.

What these names do and do not assert. What are called laws here are named conjectures under registered test. Recursion means R, rounds of a defined loop, an operational quantity of a protocol rather than a universal physical unit, and intelligence is measured only as defined capability on defined task banks. Nothing here claims the standing of a law of nature, and this registration exists so that the claim can be earned or lost by measurement and by replication the author does not run.

Twenty-two predictions are registered, each conditional on the stated scope and each severable from the others.

P1, power-law conversion and recursive-depth advantage. Within the scope, usable capability is best described by a power law in modelled recursive depth, and the depth exponent exceeds the compute-matched and token-matched breadth exponent by the registered margin. Whether it also exceeds one is P8's separate verdict.

P2, the sustainable growth profile turns over. Within the scope, the sustained growth exponent does not rise without limit as reinvestment rises, and the profile has a maximum inside the administered

range. This is a claim about the shape of that profile and is firewalled from the framework's boundary claim, which P16, P20 and P3 carry.

P3, the corrected relation upper-bounds the sustainable frontier. The maximum growth exponent that a system holds while remaining correctable lies at or below the value the ceiling relation returns from the correction-leverage exponent measured on that same system, and where that exponent is one half the value is two.

P4, correction must out-scale drift. Within the registered minimal model, relative correctable burden tends to zero where the correction exponent exceeds the drift exponent, burden holds the asymptotic advantage where the correction exponent is the smaller, and at equality the outcome, in the finite run and in the long run, is decided by coefficients, delay, backlog and saturation rather than by the exponents.

P5, the exponent is derived and not fitted. The growth exponent implied by an independently measured coupling agrees, within a registered tolerance, with the exponent fitted from the growth curve.

P6, the correction exponent is bounded for same-class correctors. Where a system corrects itself using mechanisms of its own class, the correction exponent is at or below one half.

P7, deployed correction is of the weaker class. The correction mechanisms actually used in deployed alignment practice predominantly belong to the class whose strength is fixed by construction rather than the class that can scale with the capability it corrects.

P8, measured exponents exceed the null. Across the admissible independent systems the pooled growth exponent exceeds the one-half null; an interval lying wholly below the LOWER EDGE of the registered equivalence region refutes rather than supports it, where an earlier wording said wholly below one half and so contradicted the four-outcome rule in P8's own entry, and whether it also exceeds one is the separate CLEARS UNITY verdict.

P9, the cross-domain form. Where an effective dimension can be assigned to a domain, the growth exponent follows the dimensional relation with the dimension supplied by that domain's equation of state.

P10, scorer-family dependence. A material fraction of the results this programme has scored unblinded within a single model family, ten per cent, fixed in the entry below, fails to survive re-scoring by a different family.

P11, the residual-decay and correction-capacity exponents are not practically interchangeable, and the working practice of treating them as one is an assumption rather than a fact.

P12, build order moves the exponent. The order in which a system's components are built moves the correction-leverage exponent by at least the registered minimum effect.

P13, the cross-family contrast. Where an automated alignment researcher post-trains a target model against alignment benchmarks, a researcher from a different model family than the target reduces held-out misalignment more than a researcher from the same family, at matched data and compute,

and the panels differ in measured pairwise error correlation in the direction that would explain it. The two halves are one conjunctive proposition, and an advantage arriving without the correlation difference is reported as unexplained by this framework rather than as support for it.

P14, blinded cross-family scoring. Where an automated researcher's method is re-scored by a scorer that cannot see the condition and comes from a different model family than either the researcher or its target, the alignment gain it returns is smaller than the gain the unblinded same-family scoring reported.

P15, external gains decay. Gains installed by external post-training erode under later capability-directed training that embeds no correction of its own, and more than half of what was installed is lost once the capability-training budget matches the alignment budget that installed it.

P16, the prohibition itself. A system does not hold both at once: a growth exponent above the point at which its own measured correction margin reaches zero, and a correction margin whose lower bound stays above zero, sustained across a window fixed before the run.

P17, panel corrector failures are conditionally independent on a frozen fault set. Within a corrector panel named in advance by model lineage, substrate, training provenance and scaling class, pairwise joint misses on a frozen fault set of stated type are practically equivalent to the conditional-independence baseline under the registered margin; failure-mode overlap is the outcome rather than the entry criterion, and this is the cross-sectional claim the instrument can decide rather than the temporal premise it bears on.

P18, the composition operator predicts the family. Classified in advance by how a domain composes its inputs and outputs, the scaling family that best fits that domain's data is the family the classification predicts, and not one chosen after the curve is seen.

P19, external alignment does not scale with capability. Alignment achieved by mechanisms outside a system's own recursion does not improve as the capability of the system it governs rises: its exponent is indistinguishable from zero.

P20, the ceiling relation takes the corrected form. Where a stability boundary can be located and a correction exponent measured on the same system, the boundary follows one over one minus the correction exponent rather than one over the correction exponent or a fitted constant.

P21, in-loop correction pulls away with depth. Correction that participates in each subsequent revision round outperforms correction that sees only the finished output, and the advantage widens as recursive depth rises rather than staying constant.

P22, checkable correction carries the higher exponent. The checkable correction-leverage exponent exceeds the judged correction-leverage exponent on the same systems under the same instrument, so the ceiling is strictly lower for judged correction.

Confidence is stated per prediction before any test, in the bands the Confidence Levels field fixes, except for one the author states as a sentence of his own rather than as a rung. The only unstated confidences are the two refusals, P5 and P11, where no measurement of the underlying quantity exists anywhere. The weakest link is P17 and not P6. P17 registers the cross-sectional measurement

bearing on the assumption beneath P6, and correlation among correctors refutes P17. It does not confirm P6, and no sentence in this document may be read as drawing that inference. P6 admits no supported verdict under its own rule, only no counterexample found among the number examined, and a refuted P17 stays refuted while any relation from measured dependence to the elasticity is scored on an instrument of its own.

WHAT A SUPPORTED PROPOSITION WOULD LICENSE, and what it would not, fixed here before any result exists. Each entry below states what would refute it and what falls with it. Neither says what follows if it holds, and a reader deciding whether this work is worth testing needs that as much as the refuter. It is written now, in advance, and in the restrictive form, because the danger of stating consequences after a result is that they grow: an author who has not committed beforehand to the limits of his own success will find those limits generous afterwards. Every clause below is a constraint on this programme's own future claims and none of them is an additional prediction. A supported P1 licenses the statement that recursive revision depth is a scaling variable with a measurable exponent, distinguishable from a resource-matched breadth exponent, within the registered scope. It does not license any statement about systems outside that scope, about the stability of the exponent across architectures, or about intelligence in general. A supported P4 licenses a measurable and prospective criterion for whether a system remains correctable in the tested class. It does not license the conclusion that a system with a positive margin is safe, that the specification it was given is the right one, or that faults it has not discovered are absent. A supported P5 licenses the statement that the exponent is predicted from a separately measured property rather than fitted to the curve it describes, which is the difference between a description and a mechanism. It does not license the claim that this is the only mechanism, or that the coupling behaves causally outside the range over which it was manipulated. A supported P16 licenses the statement that a boundary estimated before a run predicts the later loss of positive correction margin in the tested class. It does not license the prediction of when any particular deployed system will fail, any operational deployment decision, or the claim that a crossing is detectable while it happens. A supported P20 licenses the statement that the frontier's location follows the registered relation better than the named rivals on held-out systems. It does not license the claim that the relation is universal, nor the use of the value two anywhere the correction exponent has not been measured on the system in question. A supported P21 licenses the statement that where correction is placed changes how its advantage scales with depth, in the apparatus that varies information access. It does not license the claim that embedded correction is safe, that it removes common-mode failure, or that the engineering proposal this programme names works as an architecture. A supported P22 licenses the statement that the two typed exponents order as predicted on matched capability ladders. It does not license the claim that judged correction is unsafe, that checkable correction is sufficient, or that a safety property can be made checkable by choosing to call it so.

WHAT THE WHOLE SET WOULD LICENSE IF EVERY PROPOSITION HELD, which is the claim most at risk of growing. It would license a quantitative framework for the scaling of recursive improvement against the correction that services it, carrying a boundary that can be estimated before it is crossed and an engineering lever that moves it. It would not license the claim that a safe self-improving system can be built, and the ladder below fixes the whole list of conclusions no rung of it licenses. Nothing here is confirmatory before the scoring gate passes, as THE SCORING

INSTRUMENT IS A GATE AND NOT A CAVEAT registers in Variable Relationships. And no programme-level conclusion is drawn by counting supported propositions, on the rule stated above and again with the ladder. The twenty-two are set out in full below, each with the observation that would refute it, what falls with it, the status of the instrument that could decide it, and the author's confidence. The refuters are the substance of this registration: a proposition registered without the observation that would defeat it is half a proposition, so they are placed here, inside the registered hypotheses, rather than in an attachment.

THE EVIDENCE LADDER, AND THE ONE SENTENCE EACH RUNG PERMITS. Everything above says what a supported proposition would license. It says nothing about WHO supported it or HOW MANY TIMES, and that is the axis on which claims actually inflate. A programme's own successful run and an independent reimplementation are not the same evidence and must not share a vocabulary. Eight rungs are fixed here, before any result exists, each with the exact public wording it permits. No surface of this programme may use wording from a rung above the one a proposition has actually reached, and the wording is quoted rather than paraphrased so that the constraint is checkable by someone who has not read this document. RUNG 0, WRITTEN AND DATED: the prediction exists, dated and hashed, and nothing has been observed. Permitted where no lodgement exists: a draft prediction, written and dated, not lodged with any registry and not tested. Permitted where a lodgement exists and carries an identifier a reader can follow, which is the standing of each of the twenty-two propositions registered here: registered, with the identifier stated, and not tested.

AN EARLIER WORDING HERE READ REGISTERED, NOT TESTED, AND THAT OVERCLAIMED AT THE FLOOR OF ITS OWN LADDER. When that wording was written no unit of this programme had been submitted to any registry, so the word registered asserted a lodgement that had not then happened, and a ladder whose bottom rung overclaims cannot police the rungs above it.

REGISTRATION STATUS AND EVIDENTIAL STATUS ARE SEPARATE AXES AND ARE REPORTED SEPARATELY, because a proposition may be lodged and untested, or drafted and untested, and those are different disclosures wearing one word. Rung 0 becomes registered and not tested only when a lodgement exists and carries an identifier a reader can follow. RUNG 1, AUTHOR-RUN SUPPORT: this programme's own registered instrument returned SUPPORTED under the registered rule. Permitted: supported under the registered test, by the programme's own run. Not permitted at this rung: replicated, confirmed, established, validated, or any wording that omits whose run it was.

RUNG 2, INDEPENDENTLY REPRODUCED: a group meeting the independence conditions re-ran this programme's code on this programme's data and obtained the same numbers. Permitted: the programme's analysis reproduces. That tests arithmetic and pipeline rather than the world, and the permitted wording says so.

RUNG 3, INDEPENDENTLY REPLICATED: a group meeting the independence conditions ran the registered protocol on new data. Permitted: independently replicated, with the count of groups always stated.

RUNG 4, INDEPENDENTLY REIMPLEMENTED: a group reconstructed the instrument from the registered specification without this programme's code and without its scorer, registered its own analysis before seeing outcomes, and reached the registered verdict. Permitted: replicated across

independent implementations. This is the first rung at which the result is evidence about the world rather than about this programme's software.

RUNG 5, GENERALISED ACROSS ARCHITECTURES: rung 4 reached on systems this programme did not select, drawn from the prospectively defined eligible universe. Permitted: generalises beyond the systems the programme chose.

RUNG 6, PROSPECTIVELY PREDICTIVE: the relation, with its parameters fixed on earlier systems, predicted a quantity on a system before that system's data was opened, and the prediction landed inside the registered margin. Permitted: predicted out of sample. This rung separates a fitted description from a theory and it cannot be reached by accumulating more of the same evidence.

RUNG 7, VALIDATED WITHIN THE TESTED DOMAIN: rungs 4, 5 and 6 reached, every registered rival failed its held-out comparison, and no binding failed replication stands unanswered. Permitted: a validated quantitative theory within the tested domain, with the domain named in the same sentence.

PEER REVIEW IS METADATA ON A RUNG AND IS NEVER ITSELF A RUNG. Review can find an error at any rung and is worth having at all of them, and it substitutes for replication nowhere. The permitted form is peer-reviewed at rung N, never peer-reviewed alone.

FOUR RULES MAKE THE LADDER BINDING RATHER THAN DECORATIVE. The headline follows the highest rung reached by the proposition being described, never the highest rung reached by any proposition.

A RUNG IS ADVANCED BY A RECORD, NEVER BY AN EDIT. Promotion to any rung above zero requires the underlying entry that evidences it: a lodgement identifier for the registration axis, and for rungs two and above a checkable external study record naming the group, the protocol version it followed, whether its data was fresh, whether its implementation was its own, its declared independence, and any prior correspondence between this programme's author and the group, declared in that record, the independence conditions themselves being defined by the replication contract rather than by this registration, so that this document adds a disclosure duty and settles no eligibility question. Changing a rung field, a page or a number in a machine-readable record promotes nothing, and a record whose counts are typed rather than derived from those entries is a claim about itself. Failed, partial and inconclusive replications enter the same ledger with the same prominence, and a rung once lost is lost until it is re-earned. No programme-level claim is made by counting supported propositions. And the following do not follow from any rung of this ladder and are not claimed from these experiments at all: that alignment is solved, that any deployed system is safe, that a law holding everywhere has been found, that the framework governs anything outside its registered scopes, or any comparison of this work to a named historical figure or event.

THE DISTINCTION THE LADDER CANNOT REACH IS STATED HERE SO THAT NOBODY LATER PRETENDS IT WAS REGISTERED. Scientific establishment and historical significance are different thresholds and only the first is registrable. Rung 7 is the top of what evidence can deliver: a validated quantitative theory in a named domain. Whether that turns out to matter depends on whether the relations predict phenomena nobody registered, whether other people find them useful,

whether they survive extension, and whether they change how systems are built. None of that is in this programme's gift and none of it is preregistered here. A programme that wrote its own historical importance into its rules would have told a reader which kind of programme it is.

THE CLARIFICATION LOG, because private help is indistinguishable from private steering. Every question a replicating group asks about a protocol, and every answer given, is published in a timestamped log open to all groups. No clarification is given privately. A programme that answers one group's question in private has helped that group towards an expected answer and has no way afterwards to prove that it did not.

THE FLAGSHIP CONJUNCTION IS REGISTERED NOW SO THAT IT CANNOT BE ASSEMBLED LATER FROM WHICHEVER PROPOSITIONS HAPPENED TO SUCCEED. The chain this programme puts forward as its central result is P1, then P4, then P16, then P20, then P21: a reproducible depth relation against matched breadth; the correction-against-drift condition measured independently of it; a boundary predicted on data not used to estimate it; that boundary's location beating the registered rivals; and an intervention moving the quantity in the predicted direction. Phenomenon, law-like scaling, independent mechanism, prospective boundary, intervention. Any one of the five replicating is a result. The chain replicating is the claim, and no subset of it is ever presented as the chain.

THE ADVERSARIAL COMMITMENT, because a theory that succeeds mainly on systems its author chose has not been tested on the axis that matters. The eligible universe of systems, tasks and corrector types is defined prospectively, and a third party may select the challenge set from it without this programme's involvement. This programme commits in advance to including cases chosen to embarrass it: model families it did not select, tasks on which recursion is known to plateau, high-noise regimes, strong breadth and search baselines, corrector types outside its own taxonomy, and correction exponents away from the convenient value of one half.

AND THE SECOND-GENERATION TEST IS REGISTERED AS A COMMITMENT, NOT AS AN INSTRUMENT THAT EXISTS. Once parameters are estimated on the first systems, the next test is a sealed quantitative prediction about a system whose data has not been opened, deposited with its hash before that data is released and scored by whoever runs it. That instrument does not exist, is not claimed to exist, and reaching rung 6 without it is not possible.

Twenty-two predictions. Each is conditional on the scope above, each carries the observation that would refute it, each names what falls with it, and each states the status of the instrument that could test it. Five have no instrument at all, and they are among the largest part of the reason this registration exists: P12, P13, P14, P15 and P22, named together here and again in the amendment protocol below so that the count cannot drift between the places it appears. The sequence by which that set reached four is recorded in the predecessor development record; successor version 1.103 adds P14 as the fifth through the qualification in its entry. Three of them, P13 to P15, respond to a published apparatus (Chen, Wen and Kirchner, arXiv:2608.28945, 28 August 2026); they are prospective. That apparatus runs the kind of experiment P13 and P14 turn on, while the authors' own words, quoted in P15's entry, record that it does not test the persistence P15 is worded in. All three now have instrument status NONE, P14 through the qualification in its entry: the blinding-

evaluation draft its instrument line names does not estimate the matched gain contrast P14 requires. No proposition here carries a drafting date on its face. Nothing in this document is registered until it is submitted, at which point all twenty-two are registered together and share one date, so an internal date beside some propositions and not others would invite a reader to treat the undated ones as older or better established. Several of the later-numbered propositions are in fact among the oldest claims the programme makes, and stamping them with the day they were typed into this document would misdate the claims themselves. The drafting history is kept in the development record of this version, lodged in the originating project's storage and to be published at a persistent identifier the author assigns and links from this registration's resources, which is where it belongs. The correction histories of the entries below are recorded in the development record with it. None is a new claim: each states something this framework already asserted or already assumed, registered in the programme's study drafts but nowhere in this document. P16 and P17 came from asking what the framework's central prohibition and its weakest premise would look like if either could be supported rather than merely survive. If that apparatus is withdrawn, materially revised, or its artefacts become unavailable, P13, P14 and P15 stand as written, and P14 remains without a deciding instrument; unavailable per-item outputs also prevent evaluation, as its unchanged entry states. They are not reshaped around whatever replaces it, because a prediction rewritten to fit the experiment that arrives is no longer a prediction.

WHAT KIND OF CLAIM EACH ONE IS, because the title says predictions and they do not all predict the same kind of thing. Twelve are predictions of Laws I, II and III: P1, P2, P3, P4, P6, P8, P9, P16, P17, P18, P20 and P22, with P16 the central prohibition, P17 the registered measurement bearing on the premise P6 rests on, and bearing on it is the whole of what it does: P17 is a cross-sectional proposition about coincident failure in a panel, it parameterises one dependence-related component of the temporal accumulation model, and it does not measure or establish temporal independence, the accumulation law, the value of the correction exponent or the ARC Bound, and P22 the ordering of the ceiling by the type of fault corrected. Two are claims about the status of the derivation rather than about anything the laws govern: P5 and P11. Two are surveys of current practice, true or false whatever the laws do: P7 and P19. Two concern the scoring instrument on which every other proposition depends: P10 and P14. Four are engineering claims on which the Eden Protocol stands or falls while Laws I, II and III stand either way: P12, P13, P15 and P21, the last of them the depth interaction the protocol exists to justify. Twelve, two, two, two and four account for the twenty-two exactly. A reader assessing whether the framework survived should weigh the first group, and should not count a supported P7, P14 or P19 towards it. The rest are registered here because they were fixed in advance, and because holding them elsewhere would let them be reported selectively afterwards.

SUCCESSOR QUALIFICATION. P12, P15 and P21 remain engineering claims and under version 1.103 also carry Laws IV and V through their unchanged registers, so fifteen propositions carry registered laws and seven carry none. These are overlapping descriptions of scientific role, not twenty-five propositions or a renumbered partition. P13 stays an engineering claim in no register. The rule on whose propositions a reader weighs extends to the propositions of every registered law: a reader assessing whether Laws IV and V survived weighs P15, P12 and P21 under the verdict rules registered above, and still counts no supported P7, P14 or P19 towards any law.

P1 · POWER-LAW CONVERSION AND RECURSIVE-DEPTH ADVANTAGE. Within the scope, usable capability is best described by a power law in modelled recursive depth, and the depth exponent exceeds the compute-matched and token-matched breadth exponent by the margin registered with the deciding instrument.

BOTH EXPONENTS ARE TAKEN AGAINST ONE COMMON RESOURCE COORDINATE, and this is a condition of the claim rather than a detail of its analysis. An exponent measured in rounds of revision and an exponent measured in independent samples are not comparable numbers, and the difference between them is not an advantage. A worked case makes the failure concrete: where depth returns capability as the round count raised to six fifths while costing the square of that count, and breadth returns capability as its own budget raised to three fifths, the two look three fifths apart until the breadth arm is expressed in the depth arm's budget, at which point it returns exactly the same six fifths and the advantage vanishes. The deciding unit therefore registers one resource coordinate before any fitting, converts both arms into it, and reports the converted breadth exponent as the product of the two slopes that conversion composes. Where the two arms cannot be placed on one coordinate the incompatibility is disclosed and the contrast is reported as NOT EVALUABLE; the breadth arm is never handicapped to create a difference.

A BREADTH ARM CAN ALSO BE DEGRADED WITHOUT BEING HANDICAPPED, and the rule just stated does not reach that case. Independence at generation time does not deliver diversity of output: an arm drawing near-identical samples from one model satisfies the generation-time condition and the never-handicapped rule while returning an exponent near zero, and the depth arm then wins by the whole difference for a reason that is escape from correlated sampling error rather than conversion by depth. That is not a hypothetical here, because the compute-matched parallel exponents this document records from the programme's five earlier runs run from about minus 0.03 to about 0.31, with the fifth returning none at all. The realised diversity of the breadth arm's draws is therefore measured on the draws actually taken and reported beside the contrast, and never inferred from the generation procedure. Any floor on that measured diversity is the author's number and is recorded in the deciding instrument before any outcome is read; until it is recorded the diversity reading is reported and not scored. Independently of that floor, where the breadth arm's own exponent interval contains zero the contrast is reported as NOT EVALUABLE and never as a depth advantage, on the same terms as the incompatibility above. Whether the exponent also exceeds one is decided by P8's CLEARS UNITY verdict and not here, because a straight line is the case where the exponent equals one and sits inside the power family rather than outside it.

REFUTED BY: a registered fit in which a rival form beats the power law under the registered model comparison across a majority of estimable cells.

THE RIVAL FAMILY, NAMED IN FULL, AND THE LINEAR CASE PUT WHERE IT BELONGS. The rivals are the logarithmic form, the exponential form, a saturating form, a broken or piecewise power law with one change of regime, and the shifted finite-window solution of this framework's own growth equation. The last of those is named as a rival to this proposition and not as a friend of it: the shifted curve is what the mechanism registered under P5 actually predicts at the depths a real ladder reaches, it is not a member of the power family, and a cell won by it is credited to P5's mechanism

and never counted as a P1 FORM win. A shifted curve winning is therefore consistent with a supported P5 and with a failed P1 at the same time, and the two verdicts are reported in exactly those words. The last two are named here because a short or noisy ladder favours a single power law over both by default, so leaving them out would let this proposition be supported by the shape of its own design; they are the rivals most likely to fit a real recursive system that improves quickly and then stops. The saturating rival is not a formality: a published study of looped language models reports test-time compute scaling that follows a predictable saturating exponential decay, and it is named with its identifier in the alternative baselines. The nearest quantitative oversight literature independently uses a piecewise form with plateaus, which is the broken rival.

THIS COMPARISON HAS A MEASURED FALSE-SELECTION RATE, IT IS NOT SMALL, and the deciding unit must register its own. A numerical evaluation of the rival set named above, run before any measurement and recorded as design sensitivity rather than as evidence about any system, generated data from an exact power law and asked how often the registered comparison prefers one of the registered rivals anyway. At the span registered here and a realistic dispersion, a rival is preferred in between roughly one run in six and one run in two, varying with the exponent, the number of points on the ladder and the dispersion. Adding points lowers the rate without removing it, and a stricter information criterion does not remove it either. Where the ladder carries few points the piecewise rival wins by fitting noise with its extra parameters; where the exponent is shallow the logarithmic rival is the persistent competitor. Span alone therefore does not buy discrimination, and this document does not pretend otherwise. The consequence is registered as a condition. A FORM verdict may not be read as support until the deciding unit has demonstrated, on its own registered ladder, point count and dispersion, the rate at which its comparison selects a rival when the truth is the power law, and has registered that rate in advance. Small is fixed here rather than left to each instrument to read as it likes: the demonstrated rate of selecting a rival when the power law is true must be at most 0.05, and the demonstrated probability of selecting the power law when the power law is true must be at least 0.80, both taken from a full confusion matrix over the registered rival set and both registered before any cell is fitted.

THOSE TWO TARGETS ARE ONE REQUIREMENT READ FROM BOTH CELLS OF THAT MATRIX AND NOT TWO, and an earlier wording of this sentence offered them as two so that neither could be read as a single error rate. Selection over the registered rival set returns exactly one family, so the two probabilities are complements: a false-selection rate at or below 0.05 already forces the correct-selection probability to at least 0.95, and the second target can never be the one that is missed. Both are kept, the stricter of the two governing, and neither is offered as a second reading of the matrix. A genuinely second reading is a ceiling on the share of runs in which any one rival family is selected when the power law is true, which an aggregate rate does not capture because one persistent competitor and a spread of occasional ones return the same aggregate. That ceiling is the author's number and is recorded in the deciding instrument before any outcome is read; until it is recorded, the per-rival reading is NOT EVALUABLE and the complementary pair above governs on its own. Where either target is missed the outcome is recorded as NOT DISCRIMINATING and never as support. On the sheet it is marked inconclusive, with the target that was missed printed beside the mark. A shallow exponent and a sparse ladder each demand more points, and no ladder shorter than the span registered here is admissible in any case. This condition can only make the

requirement stricter and never weaker, and it was added because the measurement went against the design rather than for it. A straight line is not a rival family here. It is the case where the growth exponent equals one, and it sits inside the power family rather than outside it. Treating it as a separate form would let a measured exponent of one count as a power-law win, which is the reading this framework must not have: at an exponent of one, capability is proportional to depth, nothing compounds, and the conversion claim has no content. The linear case is therefore decided on the exponent, by the CLEARS UNITY verdict P8 registers, and not by the form comparison. A power-law form whose exponent does not clear one is reported as a power law that is not super-linear, in those words, and is never reported as support for SUPERLINEAR CONVERSION.

A HOLDOUT ACROSS THE WHOLE RIVAL FAMILY, WHERE THE DECIDING UNIT CAN PROVIDE ONE, AND BESIDE THE RATE ABOVE RATHER THAN IN PLACE OF IT. The false-selection targets are measured on the ladder the comparison itself is run on, so they fix how often that comparison goes wrong on data of that shape and say nothing about how well the selected form predicts levels it was not fitted to. Where the deciding unit's ladder allows it, the comparison is therefore repeated out of sample: every family in the registered rival set, the power law included, is fitted on a part of the ladder chosen before any fit and scored on the levels held out of it, and the FORM verdict is reported with each family's held-out score printed beside the in-sample comparison. Every family is held out and scored rather than the power law alone, because scoring one family out of sample against rivals fitted in sample compares a prediction with a description. The power law is preferred on this route only where it predicts the held-out levels at least as well as every other family in the set, and a power law that wins in sample and loses out of sample returns INCONCLUSIVE on FORM and is never reported as support. This requirement binds where the deciding unit can hold levels out without taking the fitted ladder below the span and the level count registered here; where it cannot, the deciding unit records that it cannot and the FORM verdict is reported as in-sample only, in those words. Like the rate above, it can make the requirement stricter and never weaker.

IT DOES NOT DISTURB P1'S OWN VERDICT, AND AN EARLIER WORDING HERE USED ONE PHRASE FOR BOTH READINGS, which let a single result be a supported P1 and support for nothing at the same time. Two labels are used from here and they are not interchangeable.

P1 SUPPORT is the registered conjunction of form and recursion, decided without reference to where the exponent sits relative to one.

SUPERLINEAR CONVERSION is the stronger reading and is decided only by P8's CLEARS UNITY verdict. A sub-linear or exactly linear power law that beats the registered functional rivals and beats matched breadth supports P1 and does not support superlinear conversion, and both halves of that sentence are the registered outcome rather than a hedge between them. The exponential rival is named because the programme's own published falsification criterion names it and an earlier draft of this proposition did not: a form faster than any power law refutes the conversion claim exactly as a slower one does, and omitting it would have left the one outcome the framework's own ceiling law is about unable to refute anything.

WHAT AN EXPONENTIAL WIN MEANS, FIXED NOW SO IT CANNOT BE READ AS SUPPORT LATER: it refutes this proposition. It is not evidence for the framework on the ground that faster growth is

more dramatic. A system whose capability grows faster than any power of depth has no finite growth exponent, so it sits above every ceiling the third law can return, and the framework's claim about such a system is that it does not stay correctable rather than that it is growing as predicted. That outcome is reported as a refutation of P1 and as an observation to be read against the ceiling law, and the scope condition excluding regimes where correction is not required to keep pace is applied before any such reading is offered. At least eight estimable cells must contribute to that majority; below eight the verdict is INCONCLUSIVE and is reported in that word, because a majority of three cells is not a majority worth the name. The floor is set here, before any fit, and it applies to support as well as to refutation: a power-law win on too few cells is INCONCLUSIVE and never support.

THE FLOOR HERE IS COUNTED IN ESTIMABLE CELLS AND THE FLOOR P8 REGISTERS IS COUNTED IN ADMISSIBLE SYSTEMS, and an earlier wording of this sentence said the two were the same count. They are not the same object. A design that crosses a handful of systems with several task families returns many more cells than systems, so one comparison can clear the count registered here and fail the count registered there, and which verdict is available would otherwise turn on which noun a scorer happened to read. Which of the two nouns governs both entries is the author's to fix, and it is recorded in the deciding instrument before any outcome is read. Until it is recorded, the interim rule registered with the admissibility conditions binds in the banking direction only.

NOT ESTABLISHED BY DEPTH GAINS ALONE: a power law in revision rounds is not by itself evidence for this proposition, because repeated independent sampling produces smooth gains in the number of samples that can be fitted by a power law without any recursion at all. This proposition is about sequential recursion, where each round consumes the artefact the previous round produced. It is therefore scored only where the depth arm is compared against a breadth arm matched on total compute and total tokens, drawing the same number of samples without passing any output back as input, on the same tasks and under the same scoring. Support requires the depth arm's exponent to exceed the breadth arm's by the margin the deciding unit registers. Where no compute-matched breadth arm exists, the result is reported as a depth-only fit and labelled as not discriminating recursion from sampling, never as support. This is the comparison the programme's own experimental paper was built to make, and stating it here stops a later depth-only result being offered for it.

NOT AVAILABLE AS A SCOPE EXIT: within the registered scope, which names a policy that revises the policy, an exponential result refutes this proposition whatever mechanism produced it. The reading that the system changed its own composition operator, so that the result is the discontinuity the programme's papers describe rather than a refutation, is not available after the result is seen. It could only have been registered in advance as a separate scope condition with its own detection rule, and it is not registered here. The papers' statement that a self-modifying system escapes the attention bound is quoted in the background as history, and it is not a licence to move this proposition's boundary afterwards.

THE LADDER REQUIRED TO TELL THE FORMS APART, registered because a short one favours the power law by default: over two or three depth levels a power law, a logarithm and a straight line are

not distinguishable at realistic scoring noise, so a comparison on a short ladder returns the power law almost whatever generated the data, and this proposition would then be supported by its own design. The comparison is scored only where the depth ladder spans the span this document requires of every exponent measurement, at least one and a half decades of recursive depth, meaning a factor of about thirty in revision rounds, across no fewer than five distinct levels. An earlier wording asked here for one decade while the admissibility conditions asked for one and a half, and the looser of the two figures sat in the proposition most exposed to a short ladder. One rule governs, the stricter: a deciding instrument may register a longer span and may not register a shorter one. The comparison is scored only where the registered sensitivity also shows that it would have selected each rival family had that family generated the data. A fit on a shorter ladder is reported as underpowered for form discrimination, with the span and the level count printed beside it, and never as support.

TWO VERDICTS HERE AND A THIRD QUESTION ANSWERED ELSEWHERE, because this proposition carries claims that are not equally well supported and must not be allowed to carry each other. FORM is whether the power law beats the logarithmic, exponential, saturating and broken-power rivals under the registered comparison on an adequate ladder. RECURSION is whether the depth arm's exponent exceeds the compute-matched breadth arm's by the registered margin. Both are reported, and this proposition is SUPPORTED only when both are. A supported FORM beside a failed RECURSION is reported in those words and means the curve is a power law that repeated sampling would also have produced, which is not what is claimed here. On the sheet it is marked inconclusive, with those words printed beside the mark. The third question, whether the exponent clears one and the conversion is therefore super-linear, is registered as P8's second verdict rather than here, so that a super-linearity failure is scored against the proposition that measures it. A reader assessing the conversion law reads all three together, and this document will report them together whatever they say. The confidence stated below governs the conjunction of FORM and RECURSION, which is this proposition's claim.

THE ALLOCATION TITRATION CONTRIBUTES EXPONENTS HERE AND CARRIES NO BOUNDARY CLAIM. Its exponents are admissible inputs wherever they meet the conditions registered here, this entry's common resource coordinate among them, and nothing it returns bears on a correction-limited boundary, as P2's entry sets out.

THE CAPABILITY SCALE CONDITION IS APPLIED TO EVERY EXPONENT OFFERED TO THIS PROPOSITION, whatever unit supplies it. The 22 August 2026 ladder-scale correction governs every unit whose exponents are offered here, and an uncorrected unit's exponents are reported on their own scale and never pooled.

FALLS WITH IT: P2 alone, and with it the integrated conversion framework in the form that treats one exponent as governing a whole ladder. The corrected map applies one principle and admits no exceptions. P2 falls outright, because a profile compared across reinvestment levels needs a form comparable between cells and has no meaning without one. P5, P8 and P16 survive the failure of a single global form and depend instead on the local elasticity being approximately constant across the window in which each is estimated: the coupling relation solves the growth equation over that window, the departure-from-null verdict needs an exponent defined per system, and the boundary

verdict needs constancy at the operating point, which the measured limits on that estimator make a real condition and not a formality. P3 and P20 are in a third position: they survive the failure of a global form, and they read the correction elasticity, which the identification statement shows is not recoverable from one observed path, so both remain NOT EVALUABLE until the crossed or off-path variation the direct boundary-mapping unit is designed to supply exists, and that is true whether P1 holds or fails. P9 is not named here and the reason is stated rather than left to inference: it is governed by the cross-domain scope and is about published measurements in other domains, whose exponents exist whether or not software recursion follows a power law. What P9 loses if P1 fails is its reading as an instance of the same conversion law; the measurement stands.

P4, P6, P7, P10, P11, P12, P13, P14, P15, P17, P18, P19, P21 and P22 do not fall with it: they are claims about correctors, about current practice and about this programme's own corpus, and each is decided by its own measurement whatever happens here. INSTRUMENT: DESIGN DRAFT, the depth titration. CONFIDENCE: high.

P2 · THE SUSTAINABLE GROWTH PROFILE HAS AN INTERIOR MAXIMUM. Within the scope, the sustained growth exponent does not rise without limit as reinvestment rises; the profile has a maximum inside the administered range.

THE TITLE STATES WHAT THE EXPERIMENT CAN ESTABLISH AND NOT MORE, which an earlier title did not. A titration over an administered range can establish that the profile turns over inside that range. It cannot by itself establish that a correction-limited boundary exists, and it cannot establish where that boundary sits, which is P16's business and P20's and P3's. An interior maximum produced by effort allocation rather than by correction still supports this proposition, because this proposition is about the shape; it is barred from supporting P3, as stated below, because P3 is about the mechanism.

THIS PROPOSITION IS FIREWALLED FROM THE FRAMEWORK'S BOUNDARY CLAIM, and the firewall is registered rather than left to a reader's restraint. The instrument that decides it is a reinvestment-allocation titration: it varies how a fixed budget divides between improving the artefact and improving the improver, and reads the growth exponent that results. That is a real and useful experiment and it is not a measurement of correction pressure. A supported P2 establishes that sustainable performance turns over inside the administered range of that dial, in the tested class. It does not establish that a correction-limited frontier exists, does not locate one, and is never cited as evidence for the ceiling relation or for the value two. Those are carried by P16, P20 and P3, decided by an instrument that measures capability growth, correctable burden and correction service on separate streams. Keeping the two apart is the difference between reporting a budget optimum and reporting a stability boundary, and this programme would rather register the distinction now than be shown it later.

RUNS THAT LOSE CORRECTABILITY STAY IN THE LEDGER, which an earlier wording did not guarantee. Every assigned run is recorded and reported, including runs whose correction margin goes negative. The sustainable maximum is estimated over the region the deciding instrument prospectively defines as correctable, and the excluded runs are counted, reported with the reason, and available for the boundary propositions to use. Dropping them from the dataset rather than

from the estimate would remove exactly the observations that carry the framework's own prohibition, and would let a profile be shaped by what was discarded.

SUSTAINED IS PART OF THE CLAIM AND NOT A QUALIFIER ADDED TO IT. The framework's own coupling relation returns an exponent that rises without bound as the coupling approaches one, so a profile of the raw exponent is predicted to be monotone, and this proposition would contradict the framework it belongs to if it were read that way. What is predicted to turn over is the exponent a system can hold while its correction still keeps pace. The profile is measured under the correctability requirement P4 registers, P4 being the proposition whose margin defines correctability and not P3, and runs that lose correctability are recorded and excluded from the profile rather than quietly dropped.

THE RIVAL EXPLANATION FOR AN INTERIOR MAXIMUM, named now so that it cannot be set aside afterwards: effort spent improving the improver is effort not spent on the object task, so a maximum can arise from allocation alone, with no ceiling of any kind involved. That rival is not hypothetical; the programme registers a separate allocation study whose own primary hypothesis is that the exponent depends on how a fixed budget is divided. The two accounts are separated by what moves the maximum. The allocation account places it where the marginal return on meta-level work falls below the marginal return on object-level work, so its location moves with that cost ratio and not with the measured correction exponent; the ceiling account predicts the reverse.

NEITHER QUANTITY IS MEASURED IN THE DECIDING UNIT, AND THAT IS DECLARED RATHER THAN CONCEALED. The unit named below measures the allocation profile alone. It fits no correction exponent, and it holds the relative price of the two kinds of work fixed by design, both calls being drawn from one round budget at one token for one token, so the cost ratio the allocation account moves with is constant there and its effect on the location of a maximum cannot be observed at all. The discriminator this paragraph states therefore cannot be operated on that unit and is NOT EVALUABLE there. What would supply it is a unit measuring the marginal cost ratio between meta-level and object-level work and the correction-leverage exponent on the same systems whose profiles it fits, so that the location of a maximum can be read against both. Until such a unit exists the discriminator is unavailable in both directions, and the consequence is stricter rather than weaker: without it an interior maximum cannot be offered in the ceiling's favour at all, and any interior maximum this instrument returns is reported as an allocation maximum and never as support for P3.

REFUTED BY: a profile that rises throughout the administered range without turning over, where the registered shape adjudication returns monotone rather than unimodal. The adjudication rule is the titration design's registered rule and is restated here so it cannot drift: an interior maximum is declared only when the bootstrap interval on the location of the maximum excludes BOTH ADMINISTERED ENDPOINTS, the lowest as well as the highest, and the registered unimodality check passes; otherwise the profile is adjudicated monotone in whichever direction the registered shape rule returns, and a maximum located at either boundary refutes.

EXCLUDING ONLY THE TOP WAS THE EARLIER RULE AND IT WAS WRONG. A profile that declines steadily from the lowest level administered has its maximum at a boundary, and it would have

satisfied a rule that asks only about the other end, so a monotone decline could have been scored as the interior optimum this proposition exists to predict. That verdict counts only from a design whose registered sensitivity demonstrates it could have found an interior maximum had one existed, since a null from an instrument that could not have detected the effect is not a null. Without that sensitivity the identical observation is inconclusive rather than refuting, and it is also the author's registered most-likely outcome, which is exactly why the distinction is fixed in advance: it must not be available to be settled afterwards, in either direction, by whoever the result happens to suit.

FALLS WITH IT: no numbered proposition. What is lost is the framework's claim that sustainable recursive performance has an interior optimum in reinvestment, and with it any reading of the titration as evidence about a boundary. An earlier wording had this proposition take P3 and the value two with it, which was a consequence of treating the titration as the boundary instrument. It is not, P3's frontier is measured elsewhere, and a monotone profile therefore leaves P3, P16 and P20 standing and testable. This is a narrowing of what this proposition carries and it is registered as one: it makes the ceiling claim harder to support through this route, not easier. INSTRUMENT: DESIGN DRAFT, the reinvestment-allocation titration. CONFIDENCE: moderate.

P3 · THE CORRECTED RELATION UPPER-BOUNDS THE SUSTAINABLE FRONTIER. The maximum growth exponent that a system holds while remaining correctable lies at or below the value the ceiling relation returns from the correction-leverage exponent measured on that same system, and where that exponent is one half the value is two.

THE VALUE TWO HERE IS THE RELATIVE RUNG. The alternative baselines record that the same three assumptions return a lower boundary, two thirds at these dials, once the question is absolute burden over a named exposure rather than the ratio of correction to burden. This proposition bounds the relative rung, because the correctability it conditions on is the margin P4 registers. Nothing in this proposition is offered against the absolute rung, and the value two is never to be quoted without the rung it belongs to.

THE TITLE IS AN UPPER BOUND AND NOT AN EQUALITY, which two earlier titles obscured in different ways. The registered test is one-sided: it asks whether the sustainable frontier lies at or below the value the relation returns. A frontier well below that value satisfies this proposition without the relation having located anything, and the first title, saying the ceiling takes the derived value, invited a reader to expect an equality the test does not examine. The second, the derived value upper-bounds the sustainable maximum, was one-sided but still spoke of a derived value as though the number were fixed, when what the relation returns depends on a measurement nobody has made. The title now names the relation and what it does. Whether the boundary follows this relation rather than a rival is P20's question, and P20 is the calibration test on which a supported verdict here depends. The correctability clause is in the statement and not only in the refuter, because the conjunction is the claim: this framework does not say a system cannot reach a higher exponent, it says a system cannot hold one and stay correctable.

FIVE OUTCOMES, because a one-sided test has an informative middle that an earlier wording had nowhere to put. REFUTED, on a replicated exceedance of the returned value while correctability remains demonstrably positive, under the conditions in the refuter below. SUPPORTED, only where

two things hold together: P20's held-out form test passes, so the relation is the one the frontier follows rather than a rival that happens to sit nearby, and the paired difference below falls within the registered tolerance. INCONCLUSIVE in every case the outcomes named here and below do not reach, and the case named expressly is the run in which the correction-leverage exponent has been measured, the design's own sensitivity is adequate, no exceedance is found, the paired difference falls within the registered tolerance, and P20's held-out form test fails: it is named now rather than settled once a result is in view, and it reaches the scoring sheet marked inconclusive with the failed form test printed beside the mark.

THAT TOLERANCE IS NOT YET A NUMBER, and the supported outcome cannot be returned until it is. The equivalence condition in the measurement field names margins for three other propositions and none for this one. The tolerance is the author's to fix, it is recorded in the deciding instrument before any pair of estimates exists, and until then this outcome is NOT EVALUABLE, which is the status this entry already carries for the unmeasured exponent it depends on. CONSISTENT, NOT SUPPORTIVE, where a design whose own sensitivity shows it could have detected an exceedance of the registered size finds none, but the frontier is not located near the returned value: that is consistency with an upper bound and it is not evidence that the bound is where the relation says, and it is reported in those words and tabulated as inconclusive.

NOT EVALUABLE, where the correction-leverage exponent has not been measured, which is the present state.

THE COMPARISON IS PAIRED AND ITS UNCERTAINTY IS JOINT, which is not the same as two intervals sitting near each other. The quantity scored is the difference between the frontier exponent measured on a system and the value the relation returns from that same system's correction-leverage exponent. Both inputs are estimated, both carry error, and the second is a non-linear function of an estimate, so its error is neither symmetric nor independent of the first. The difference is therefore estimated with its own interval by the joint bootstrap or measurement model the deciding unit freezes in advance, propagating the covariance of the two estimates and the curvature of the relation. Comparing two separately computed intervals and reporting agreement when they overlap is not admissible here, and neither is treating the correction-leverage exponent as a fixed known input to the relation when it is a fitted quantity with an interval of its own.

THE RIVAL THE RELATION MUST BEAT IS NOT ONLY THE RETRACTED FORM. A more general model lets the correctable burden generated per unit of capability change with capability rather than staying fixed, and returns a frontier of one over the quantity one plus that burden-intensity elasticity minus the correction-leverage exponent. The registered relation is the special case where that elasticity is zero. Where the deciding unit estimates it, the general form is reported beside the registered one, and a supported verdict here requires the elasticity to be practically equivalent to zero under the registered margin and the simpler model to predict held-out frontiers at least as well. This rival is registered now, before any measurement, because it is the first thing a reader who works through the derivation will propose, and because it would otherwise be available afterwards as an explanation for a frontier that sits somewhere the registered relation did not predict.

REFUTED BY: an interval on the profile maximum lying entirely above the derived value, which replicates on fresh data at the same settings, and which is obtained while the system is still correctable, meaning its correction exponent out-scales its drift exponent with the lower bound of that margin above zero across a window fixed before the run.

WHY THE SECOND HALF IS NOT OPTIONAL: this proposition is about what a system can sustain, not about what it can momentarily reach. A criterion written on magnitude alone would fire on the transient excursion the framework itself predicts, and would convert a confirmation into a refutation. The programme's published alignment paper already carries this conjunct in its own criterion, and a registration weaker than the paper it registers would be worth nothing. An excursion above the derived value in a run that has already lost correctability is reported in full, and is scored as consistent with this proposition rather than against it.

NOT EVALUABLE UNTIL: the correction exponent is measured. The derived value is whatever the ceiling relation returns from a measured correction exponent, and that exponent has never been measured, by this programme or by anyone. Until the correction-exponent estimation reports, there is no derived value for a maximum to be compared against, and this proposition returns NOT EVALUABLE rather than either verdict. Reading the printed value two as the derived value in the meantime would assume the quantity the proposition is about, and is excluded here in advance.

WHAT WOULD MAKE IT EVALUABLE, put as a number so that the wait is bounded rather than open: the ceiling relation's sensitivity to the correction exponent at one half is four, since the derivative of one over one minus the exponent is one over the square of one minus the exponent. A half-width of 0.08 on the correction exponent, which is the figure the feeding unit registers, therefore carries a half-width of about 0.32 on the derived value, which cannot separate two from two and a half. Resolving the derived value to about one tenth requires the correction exponent to about 0.025, roughly a threefold narrowing of that unit's present interval. That is the target this proposition sets for the instrument, and until it is met the value two is not presented as a result anywhere in the programme, in this registration or outside it.

SAME-SYSTEM PAIRING IS REQUIRED, and it is required here rather than left to the instruments: the profile maximum and the correction exponent must be measured on the same system, in the same composition class, at the same capability level. Taking the maximum from one unit's titration and the correction exponent from another unit's roster compares two quantities that were never joined, which is the defect a separate drafted unit exists to detect one level up, reproduced here. A pairing drawn from different systems is reported as a roster-level comparison, labelled as not deciding this proposition, and never scored as support or refutation.

FALLS WITH IT: the headline number. P1, P2 and P4 survive; the measured ceiling replaces the derived one everywhere in the programme. INSTRUMENT: DESIGN DRAFT, the paired frontier-and-exponent unit, which measures a frontier and the correction-leverage exponent on the same system so that the two enter the comparison above as a pair. It is named for that pairing rather than by a stem it would otherwise share with the direct boundary-mapping unit named in P16's and P20's entries, which is a different object: that unit reads a located zero crossing on held-out confirmatory cells, and one word between two names is not a distinction a scorer can be asked to hold. CONFIDENCE: low, because

support requires the untested marginal-burden premise, same-system pairing, a target-axis correction elasticity measured far more tightly than anything now available, and P20's held-out discrimination against the rivals registered beside it. An earlier wording gave the reason as inheritance from P6, which is wrong: P3 reads the exponent measured on the system in front of it and does not require P6's universal bound.

P4 · CORRECTION OUT-SCALING DRIFT IS WHAT CARRIES THE ASYMPTOTIC ADVANTAGE. Within the registered minimal model, relative correctable burden tends to zero where the correction exponent exceeds the drift exponent; burden holds the asymptotic advantage where the correction exponent is the smaller; and at equality the outcome, in the finite run and in the long run, is decided by coefficients, delay, backlog and saturation rather than by the exponents. An earlier wording said conformance persists only while corrective capacity grows faster than drift, which asserts that strict inequality is necessary, while this proposition's own decision rule has always treated equality as undecided rather than refuting. The three regimes are registered here so that the statement and the rule say the same thing.

AN EXPONENT ORDERING IS NOT A LOSS PREDICTION, AND THE MODEL CONNECTING THEM IS REGISTERED SEPARATELY OR NOT AT ALL. Measured service and measured burden give the sign of a margin. What a given margin implies for how much conformance is actually lost, and when, needs a response model with its own coefficients, its own initial state, its own horizon and its own uncertainty, and this proposition supplies none. Any unit converting a margin into a predicted loss registers that model before it runs, and an inverse-time quantity taken from this programme's earlier work is not automatically a repair-throughput quantity: the two share a dimension and not a definition. Where no response model is registered, the outcome of this proposition is a statement about the ordering of two exponents and is reported in exactly those words.

WHOSE INTENT, ANSWERED IN THE STATEMENT RATHER THAN A FOOTNOTE. What the deciding instruments score is conformance to a specification fixed and recorded before the run, under a frozen rubric, and not the system's own self-description. Nothing here is a measurement of a machine's endogenous intent, and no claim about one is made or implied.

NECESSARY UNDER THE MODEL, NOT SUFFICIENT FOR SAFETY, and the difference is registered before any result. What the deciding instrument tests is a statement inside a stated dynamical model in which correctable burden accumulates and correction services it: within that model, a correction exponent below the drift exponent produces eventual divergence unless coefficients, delays or a change of regime intervene.

THE EARLIER STATEMENT SAID ONLY WHILE, and that was too strong at the boundary, which is corrected here rather than left for a referee. Three regimes exist under the model and the statement above names two of them. Where the correction exponent is below the drift exponent, burden eventually outruns correction. Where it is above, correction holds an asymptotic advantage. Where the two are equal the ratio of correction to burden is constant, and a constant ratio is not a ratio of one: if the coefficient on correction exceeds the coefficient on burden, and delay and initial backlog are controlled, conformance can persist indefinitely without the exponent being strictly greater.

Strict inequality is therefore sufficient for an asymptotic advantage and is not necessary for conformance to persist. No verdict of refutation may be returned against this proposition on the strength of a system holding conformance at equality of the exponents, and the deciding instrument reports the coefficients and the backlog alongside the exponents so that the equality case can be read rather than assumed to have failed. A positive margin does not establish that the specification is the right one, that every consequential fault is discoverable, that discovery and service delays are short enough, or that failures do not arrive together. A supported P4 is therefore a scaling condition and never a safety certificate, and this document will not describe it as one.

REFUTED BY: an observed system that remains aligned on the registered measure while its measured correction exponent sits below its measured drift exponent, over a horizon long enough for the shortfall to have shown.

THE REGIME MUST BE ONE IN WHICH THIS PROPOSITION CAN FIRE, and where it is not, nothing is banked. At the growth exponent this programme's surviving cross-architecture estimate reports, the burden model implies a drift exponent near minus one, so any non-negative correction exponent clears it and the margin is positive by construction, while the corrected ceiling relation returns at least one across the whole admissible range, so no cell sits near a boundary. A positive margin read in that regime is arithmetic rather than evidence. The deciding unit therefore reports the realised range of the growth exponent beside the verdict, and where no cell in a run reaches a growth exponent above one this proposition returns NOT EVALUABLE rather than reporting a margin positive by construction as support. That is a restriction on what may be banked and not a change to the refuter, which stands exactly as written.

THE HORIZON IS PART OF THE REFUTER, because the claim is asymptotic and a run is not: a system whose correction exponent sits below its drift exponent stays aligned for some time whatever the exponents say, because divergence has not yet had room to appear, so a short run would refute this proposition by being short. The refuter fires only where the run reaches the horizon registered with the deciding instrument, fixed as the depth at which the projected divergence under the measured exponents exceeds the registered alignment tolerance. A run ending before that depth is reported as inconclusive, with the depth reached and the depth required printed beside it.

WHAT IS MEASURED IS THE MARGIN AND NOT THE SPEED: the binding quantity is the difference between the two exponents rather than the rate at which the system improves, so a fast system with a positive margin is not evidence against this proposition and a slow system with a negative margin is not evidence for it. A separate drafted unit sharpens this into a threshold, that the drift slope falls monotonically with the correction ratio and that a critical ratio exists. That sharpening is not registered here as a claim: if it reports, it strengthens this proposition without having been promised by it.

FALLS WITH IT: no other prediction outright, but the operational meaning of the words sustainable and correctable in P2, P3 and P16 goes with it, and P22's ceiling consequence is unscoped alongside them, because each of those is defined by the margin registered here. A proposition whose regime has lost its definition is unscoped rather than upheld. A scorer meeting this case marks P2, P3, P16 and P22's ceiling consequence untested, never held.

AN EARLIER WORDING OF THIS LINE NAMED ONLY P2 AND P3 while the scoring sheet, the note beneath it, the dependency map and P16's own dependency line all named the wider set, so a scorer who read this entry alone would have left P16 and P22's ceiling consequence standing on a failed P4. The enumerations are made one here. A CHECK ON THE BURDEN MODEL that costs nothing extra and runs before any boundary is read. The alternative baselines record that the ceiling relation is this criterion with the drift exponent fixed by the burden model rather than measured, which makes the burden model testable on its own. The deciding unit measures the drift exponent directly and compares it with the value the burden model implies for the growth exponent observed on the same system, which for a power law is that exponent less one, divided by itself. Agreement within the registered exponent-scale margin leaves the ceiling standing. A material disagreement refutes the burden model rather than the boundary written on top of it, and is reported as such: it would mean the two propositions this programme treats as one condition are not one condition on that system. The check is registered because a framework that derives a boundary from an unmeasured intermediate quantity should measure that quantity first.

THE SAME CHECK IS WRITTEN IN TWO COORDINATES, AND THE MAPPING BETWEEN THEM IS REGISTERED HERE SO THAT IT IS SCORED ONCE. The comparison above reads the drift exponent directly against the value the burden model implies. The direct boundary-mapping unit reads the same quantity in another coordinate, as the burden-intensity elasticity, which is the rate at which the correctable burden generated per unit of capability changes with capability, and it tests that elasticity against an equivalence region at zero. The burden model registered here is the case in which that elasticity vanishes, so the two are one check and not two: an elasticity practically equivalent to zero under the registered margin is the same finding as agreement between the measured drift exponent and the implied value, and an elasticity whose interval lies wholly outside that region is the same finding as the material disagreement above. The burden-intensity generalisation that P20 carries among its named rivals is that same quantity under a third description. The check is therefore scored once, on whichever coordinate the deciding unit reports, and the result is carried to the other two entries with its coordinate stated rather than scored again there. Registering the mapping in advance stops one result being counted twice, once as a check discharged here and once as a rival defeated there. INSTRUMENT: DESIGN DRAFT, the stability unit, extended with the direct drift measurement and the comparison above. CONFIDENCE: moderate.

P5 · THE EXPONENT IS DERIVED, NOT FITTED. The growth exponent implied by an independently measured coupling agrees, within a registered tolerance, with the exponent fitted from the growth curve.

REFUTED BY: systematic disagreement beyond tolerance across a majority of estimable systems.

A MAJORITY OF WHAT, AND THE FLOOR IS NOT YET A NUMBER. This document rules elsewhere, for the identical reason, that a majority of three cells is not a majority worth the name, and no minimum count of systems is registered here, while the deciding unit scopes its own claim to a held-out panel of at least three systems from three lineages. The floor is the author's to fix. It is recorded in the deciding instrument before any outcome is read, and until it is, this proposition is NOT EVALUABLE and no majority verdict may be returned on it. The floor binds the supporting side and

not refutation, and that asymmetry is registered here rather than carried by the deciding unit alone: a bare majority of a three-system panel is two, and two systems are not a majority worth the name for this proposition, so a supported verdict is never returned on a bare majority alone, while a refuting majority is not held to the same bar, because raising the bar for refutation would weaken the refuter this proposition exists to expose the framework to. The floor is not lowered to reach a verdict. This document's own system floor, for the proposition that tests measured exponents against the null under the scope that governs this one, is eight admissible systems; the deciding unit's held-out panel of three sits below that figure, which is stated here rather than elided, and a supported result there is reported as support on three lineages, scoped to them, and never as though it had met that floor.

AND THE DECIDING UNIT'S OWN CONCESSION IS REGISTERED HERE, BESIDE THE PAIR IT QUALIFIES. That unit concedes in its own text that a change of regime between the panel the coupling was measured on and the panel it predicts is not refuted by its design, and that a supported result is reported as consistent with such a change rather than as excluding it. The change of regime is named here rather than left as a word: the coupling measured on the titration panel may fail to govern the held-out panel because the two panels differ in some way the eligible universe does not control, and a failed comparison cannot distinguish that from a false framework. It is mitigated only by drawing both panels from one prospectively defined eligible universe and by assigning panel membership at random before any run, and neither the seal nor the random assignment separates a framework that is right from two panels that share something that universe does not control. The pair verdict above is read under that limit: SUPPORTED with IDENTIFIED means the sealed prediction held and the two measurement routes agreed on the panels run, and it does not mean that the same coupling would predict a panel in a different regime.

TWO RESULTS ARE SCORED HERE AND THEY ARE REPORTED SIDE BY SIDE, because this proposition carries a prediction and a mechanism and they can come apart. The PREDICTION result asks whether the exponent implied by the sealed coupling agrees with the exponent the held-out trajectory returns, and takes the outcomes SUPPORTED, REFUTED and INCONCLUSIVE. The IDENTIFICATION result asks whether the coupling was measured off the path it predicts, and takes IDENTIFIED, NOT IDENTIFIED and INCONCLUSIVE according to the agreement of the two registered measurement routes within their registered margin. A disagreement between the routes never converts a correct prediction into an inconclusive one: the case in which a nuisance rate that grows with capability drives both the measured coupling and the trajectory returns a coupling that predicts the path in every run while the mechanism reading is wrong, and a rule that returned INCONCLUSIVE for the whole proposition would throw away a correct prediction, while one that returned SUPPORTED without the route check would credit the wrong mechanism. The proposition-level line is therefore the pair, in both words, and never a single one of them.

ON THE SCORING SHEET THE MARK FOLLOWS THE PREDICTION RESULT AND THE IDENTIFICATION WORD IS PRINTED BESIDE IT, because neither identification word is among the four the sheet admits and this entry forbids reducing the pair to one of them. A supported prediction returned with NOT IDENTIFIED therefore reaches the sheet marked supported with that word printed beside the mark, and is never read as establishing the mechanism; an identification

returning INCONCLUSIVE is printed beside the mark on the same rule. The sheet carries both halves of the pair, and the mark alone is never the verdict.

WHY THIS PROPOSITION CARRIES THE FRAMEWORK'S CLAIM TO PREDICT, and why the algebra alone carries nothing. The relation that gives the growth exponent as one over one minus the coupling is the ASYMPTOTIC form of the solution of the growth equation the framework assumes, and an earlier wording here called it the closed-form solution, which hides an integration constant that matters at every depth a real ladder reaches. Integrating that equation from a positive starting capability gives capability at depth as the starting capability raised to one minus the coupling, plus a term linear in depth, the whole raised to the reciprocal of one minus the coupling. That is a shifted power and not a pure one, and it approaches the pure power only once the starting term becomes negligible.

THE GAP IS NOT SMALL AT THE DEPTHS THIS PROGRAMME REGISTERS. With a coupling of one half, a unit starting capability and a rate of one tenth, the exact solution is the square of ninety-five hundredths plus one twentieth of the depth. Its asymptotic exponent is two, and its two-point slope in logarithms between the first depth and the thirty-second is about fifty-four hundredths. A comparison that set an implied two against a fitted fifty-four hundredths would report a refutation of the very equation those numbers came from, which is a defect in the comparison and not a finding about any system.

P5 THEREFORE COMPARES THE IMPLIED EXPONENT AGAINST THE FINITE-WINDOW EXPONENT THE COMPLETE SOLUTION PREDICTS, over the registered depth window and from the registered starting capability, and the deciding instrument implements that comparison and no other. Comparing against the asymptotic exponent instead is admissible only where the instrument registers, before any run, the condition under which the starting term is negligible across its own window. The two quantities are therefore one-to-one by construction: given either, the other follows without any measurement. It follows that fitting the relation to solutions of its own equation, or reading a coupling off the growth curve the coupling is then said to predict, verifies arithmetic and establishes nothing about the world. A demonstration that the closed form satisfies the equation it solves is an internal consistency check and is reported as one, never as validation of the law. What would make it empirical is stated as the whole content of this proposition: the coupling is estimated by a manipulation, on records disjoint from the growth trajectories it is then compared against; the exponent that estimate implies is written down and sealed before those trajectories are opened; and the sealed prediction is compared with the fitted exponent under the equivalence margin condition D registers, which for this comparison is 0.10 on the exponent scale, the margin this document holds in common for every exponent comparison it decides. This is the framework's own strongest unexploited falsifier and it is also the single result that would convert a fitted scaling curve into a mechanistic prediction, which is why this entry was written with no instrument named rather than held back until one existed, and why the design that arrived afterwards is recorded below as the one this entry specified rather than as a new idea.

FALLS WITH IT: the claim that the framework predicts rather than describes. The measurements survive; the derivation does not. IT CARRIED NONE UNTIL THAT UNIT WAS WRITTEN AND THE CHANGE IS RECORDED RATHER THAN ABSORBED. This entry carried the words no instrument from the day

it was written, with the note that building one later could not then be presented as a new idea. That note did its work: the design now drafted is the one this entry specified, it measures the coupling by manipulating the retained fraction of a system's own output at a fixed round on systems disjoint from those it predicts, it hash-commits the implied exponent before the sealed windows open, and it compares against the finite-window exponent of the complete solution rather than against its asymptote.

WHAT THE DRAFT ALREADY ESTABLISHES, AND IT IS NOT A RESULT ABOUT ANY SYSTEM. Its sizing was measured before any collection was proposed, and it found that the coupling is the easy part: at the registered configuration the coupling is estimated tightly enough that its error moves the predicted exponent by less than a tenth of the margin, while the nuisance rate and the scored capability carry the rest. Its design-sensitivity rates are a separate matter, and what this entry said about them is corrected here: an earlier wording summarised them as measured at the registered configuration and they were not. They were measured on 5 September 2026 on a single-ladder approximation to that design, whose titration panel moves the retained fraction alone and carries no state factor, no second route and no negative-control cells, so it cannot return the identification result at all; whose sealed window stops at a depth shorter than the registered one, where the margin resolves a coarser shift in the coupling; and whose held-out panel is five systems where the unit registers three. Those rates therefore bound what a design of that shape could resolve under a stipulated noise model, and they are not this instrument's operating characteristic. The binding requirement is a capability measurement precise to about two per cent per observation, which is a demand on the scoring instrument rather than on the systems, and it makes this unit NOT EVALUABLE until the scorer-validation units have passed. That ordering is a numerical fact about this design and not a preference.

WHAT AN ADMISSIBLE INSTRUMENT WOULD HAVE TO DO, restated beside the proposition so that it is a specification rather than an aspiration: the criterion is registered in the variable specification above and is repeated here, because a status word alone, whether the NONE this entry once carried or the DESIGN DRAFT it carries now, tells a reader nothing about what would satisfy this proposition. The coupling must be recovered from a quantity that is not the growth curve this proposition compares it against, and any design meeting the three registered criteria of identifiability, provenance separated from quality, and non-circularity is admissible.

AN EARLIER WORDING MADE ONE DESIGN MANDATORY HERE WHILE THE VARIABLE SPECIFICATION ABOVE CALLED THE SAME DESIGN AN EXAMPLE, so a later and better-identified instrument would have been inadmissible for failing to instantiate an illustration. One admissible family, named as an example and not as the requirement, varies the retained fraction of its own output that each round may consume, across levels and across depths, so that the estimate does not inherit the fit it is being tested against. An instrument that reads the coupling off the same curve satisfies nothing here, however well it fits. INSTRUMENT: DESIGN DRAFT, the coupling-identification titration, written and dated as its own unit, not registered and not run. CONFIDENCE: not stated, because the author has no measurement to be confident about.

P6 · THE CORRECTION EXPONENT IS BOUNDED FOR SAME-CLASS CORRECTORS.
Where a system corrects itself using mechanisms of its own class, the correction exponent is at or below one half.

REFUTED BY: a same-class corrector whose measured correction exponent interval lies entirely above one half.

MINIMUM DETECTABLE EXCEEDANCE, registered because the refuter cannot fire below it: refutation asks for an interval lying entirely above one half, so at the registered feeder precision it can only fire when the true exponent sits far enough above one half for the whole interval to clear it. The correction-exponent unit registers a 95 per cent half-width of about 0.06 to 0.08, worst case 0.108. A same-class corrector genuinely at 0.55 therefore returns an interval of roughly 0.48 to 0.62, which straddles one half and refutes nothing. At that precision this proposition is unrefutable anywhere between one half and about 0.58, and up to about 0.61 in the worst case. That band is stated here rather than discovered later: a measured exponent inside it is reported as exceeding one half at insufficient precision to refute, with its interval printed, and is never reported as consistent with the bound. Narrowing the band is a matter for the feeder's per-cell floor, recorded against that unit, and not for a margin chosen here after the fact.

THE BAND IS A STATEMENT ABOUT AN INSTRUMENT AND NOT ABOUT THE WORLD, and it is registered as a requirement rather than offered as an excuse: no survey cell counts towards this proposition unless the unit supplying it has demonstrated in advance, on its own ladder, the probability with which it would detect an exceedance of the size that matters, and that probability is printed beside the cell. A cell that cannot clear the registered detection probability is recorded as examined and uninformative, which is a different outcome from a cell that looked and found nothing.

WHAT THIS PROGRAMME'S OWN MODEL PREDICTS THE EXPONENT TO BE, put beside the band above because the two are only informative together. The measurement model from which one half is derived returns one half only where a corrector's channels are uncorrelated. At any positive correlation the elasticity falls below one half, and it falls further as the panel grows, to about 0.06 at a correlation of one tenth over sixty-four channels. If that model is right, a real same-class corrector sits far below one half and further still below the exceedance the refuter needs, so NO COUNTEREXAMPLE FOUND AMONG N is what the model predicts whether or not the bound is true, and a long run of it is the design working rather than the bound holding. That reading is fixed here so that it cannot be chosen later: a null run of this kind is evidence about the model and is never evidence for the bound, and it is reported in those words with N stated beside it. Nothing in this paragraph moves a threshold, withdraws a verdict or adds one, and the supportable population form named below stays outside this registration.

NOT SUPPORTED BY ITS ABSENCE: one counterexample refutes this proposition, and finding none does not establish it. A survey that examines a handful of same-class correctors and reports no exceedance is reported as no counterexample found at that sample, with the number examined stated beside it, and never as support for the bound.

THIS PROPOSITION IS UNIVERSAL IN FORM AND ITS VERDICTS ARE FIXED ACCORDINGLY, because no finite survey establishes a universal bound over an open class. Three verdicts are available and support is not among them: REFUTED by a single valid counterexample; NO COUNTEREXAMPLE FOUND AMONG N, with N stated, which is the realistic long-run state, is never reported as support, and reaches the scoring sheet marked inconclusive with the number examined printed beside the mark, because a null run at the registered feeder precision is what the model recorded above predicts whether or not the bound is true and so decides nothing about the bound in either direction; and INCONCLUSIVE where the survey lacked the precision to have found a counterexample of the registered size. An earlier wording offered support after twelve same-class correctors all fell at or below one half. Twelve is a useful survey and it is not a universal bound, so that offer is withdrawn. If the author later wishes a supportable form, it must be registered separately as a population claim, an upper bound on the proportion or the quantile of same-class correctors exceeding one half, which a finite sample can decide.

THE POPULATION AND THE LIKELY LONG-RUN STATE, both registered so that neither is settled afterwards: the class of same-class correctors is the one fixed by the programme's corrector taxonomy, cited by name and version in the deciding instrument's own registration rather than described afresh here, because a population redescribed in two places drifts between them.

THAT TAXONOMY MUST BE FROZEN BEFORE THIS PROPOSITION CAN BE SCORED, and it is not frozen today. A universal claim whose population is settled after the survey is not a universal claim; it is a description of whichever correctors were convenient. The taxonomy is therefore versioned and hash-pinned before any corrector is measured against this proposition, and until it is, this proposition is NOT EVALUABLE on the same ground as the other propositions in this document whose instruments are blocked on a rule that is not yet settled. Scaling class, lineage, substrate, training provenance, access to external ground truth and failure-mode dependence are separate axes in that taxonomy and none of them is a proxy for another. And the honest expectation is recorded now: the support bar of twelve correctors with every interval at or below one half is unlikely to be met even if the bound holds, so the realistic long-run state of this proposition is no counterexample found at the number examined. That is not support, this document has already said so, and saying it twice is cheaper than letting a long run of null results accumulate into a claim nobody registered.

WHAT ONE HALF DEPENDS ON. The value is not a consequence of independence alone: it follows from the four conditions stated with the alternative baselines above. A measured exceedance refutes this proposition as written; a measured shortfall is consistent with it and points at which condition failed, and the deciding instrument reports the correlation it observed alongside the exponent. None of this softens the kill condition below, which is unchanged.

FALLS WITH IT: the at-most-two reading for same-class correctors, and no numbered proposition. P3 does not fall with it. P3 bounds the sustainable frontier by the value the relation returns from the exponent measured on that same system, so an exponent above one half raises the value P3 is tested against rather than removing the test.

THE BOUND IS FIXED ON ONE CAPABILITY SCALE AND ON NO OTHER. A numerical exponent is a property of the scale on which capability is expressed, so the value one half is meaningful only

against a ratio scale justified independently and fixed before any outcome is seen. This bound is registered on the scale that admissibility condition B fixes. A convenient benchmark transform, a latent score or a rescaling chosen afterwards would change the number without changing anything about the system, and a bound stated on such a scale would not be the bound stated here.

WHICH AXIS AN AUDIT FREEZES, AND WHAT WOULD SUPPLY THE REST, said so that the freezing of one axis is not read as the unblocking of this proposition. The census of deployed correction named at P7 partitions correctors on the scaling axis, what a mechanism's strength may depend on. The two axes this bound actually rests on are access to external ground truth and failure-mode dependence, and a scaling partition settles neither. Freezing the scaling axis therefore removes one part of the block and leaves the population unfrozen on the two that carry the bound; what would supply them is a corrector-classification unit coding external grounding and failure-mode overlap on the same mechanisms, versioned and hash-pinned on the same rule before any corrector is measured against this proposition. Nothing in the proposition moves here, and what changes is the account of how much of the block one unit removes.

THAT ESTIMATOR CARRIES ITS OWN REGISTERED DIRECTIONAL SECONDARY, named here so that the deciding unit's stated expectation is visible where this proposition is scored. The secondary conjectures the value this proposition asserts, an exponent at or below one half, for correctors that share their system's substrate and fall in the scaling class that unit admits to estimation, on the stated reasoning that corrections aggregated by a corrector sharing its system's substrate cannot be anti-correlated with themselves; it is scored per corrector family on three labels, SUPPORTED where the blind interval lies wholly at or below one half, CHALLENGED where it lies wholly above, and SPANNING otherwise. Its CHALLENGED condition and this proposition's refuter test the same inequality at two different grains, the family and the single corrector, so a challenged family is evidence in the refuting direction and becomes the counterexample this proposition asks for only where the interval is read on a corrector the frozen taxonomy admits as same-class. A supported family is not support for this proposition, whose universal form admits none; it is one family in which no counterexample was found, which is the verdict this entry already names as NO COUNTEREXAMPLE FOUND AMONG N. The estimator's primary remains an estimate with no predicted value, and naming its secondary here does not promote it.

The assumption beneath this bound is attacked separately by P17, which measures a cross-sectional proxy for it rather than the assumption itself, and a drafted instrument measures that proxy. This proposition is not the framework's weakest link, though an earlier version of this document described it as such. P17 is. The two move oppositely: correlation among correctors refutes P17 while driving the exponent further below this bound, which is consistency with it and never confirmation, since this proposition admits no supported verdict. What correlated failure removes is the value two, not this proposition, and it leaves a ceiling nearer one than two.

INSTRUMENT: DESIGN DRAFT, BLOCKED ON A FROZEN CORRECTOR TAXONOMY, the correction-exponent estimator. The estimator is drafted; the population this proposition is universal over is not yet frozen, and the status names that rather than implying a runnable draft. CONFIDENCE: low.

P7 · DEPLOYED CORRECTION IS OF THE WEAKER CLASS. The correction mechanisms actually used in deployed alignment practice predominantly belong to the class whose strength is fixed by construction rather than the class that can scale with the capability it corrects.

REFUTED BY: a preregistered classification in which the scaling class predominates.

THE CODEBOOK IS FIXED BEFORE ANY MECHANISM IS READ, and mixed is a verdict rather than a rounding. Deployed correction is not built to this document's taxonomy, and several current mechanisms sit across it: a model critiquing its own output against a written constitution scales with the model and is fixed by the constitution; deliberative approaches place a fixed specification inside a scaling process. A binary forced onto that population would return whichever answer the coder preferred. The audit therefore codes each mechanism on both registered axes, its scaling class and its independence from the generator, records placement, update cadence and external grounding as covariates, and carries an explicit mixed or ambiguous category which is counted and reported rather than allocated. Two coders work blind to which class favours the framework, their agreement is reported, and a tie-break rule is registered in advance. A predominance verdict is available only over mechanisms the codebook classifies cleanly, with the mixed count printed beside it; where the mixed category is the largest, that is the finding and it is reported as one.

THE DENOMINATOR AND THE THRESHOLD, FIXED BEFORE THE AUDIT RUNS, because predominantly without one is not an observation: the population is the correction mechanisms in documented use at named frontier deployments within a window stated in the audit's own registration. Two counts are taken and both are reported. The first counts distinct techniques once each. The second weights each technique by the number of deployments using it. Predominance is more than half on a count, and the verdict is SUPPORTED only when both counts agree, REFUTED only when both agree the other way, and INCONCLUSIVE when they disagree, because a result that turns on which way the population was counted is not a finding about deployed practice.

WHAT A SUPPORTED VERDICT HERE DOES NOT ESTABLISH, registered before the audit runs because it cuts against this programme's own argument: this proposition is about the scaling axis alone. A mechanism fixed by construction may be, and a formal verifier probably is, among the most independent correctors available, and independence is the axis the ceiling actually depends on. So a finding that deployed practice predominantly uses the fixed class does not by itself show that deployed practice is unsafe; it may be pointing at mechanisms that are weaker on one axis and stronger on the other. Any claim from this proposition to a safety conclusion must carry the independence measurement as well, and this registration does not license the shorter argument.

FALLS WITH IT: the programme's argument that current practice cannot keep pace, in the form that argues from scaling class alone. The framework itself is untouched.

THIS IS A DATED CENSUS AND IT IS REGISTERED AS ONE. What deployed practice does is a fact about a date rather than a law, and practice can change without anything in this framework being wrong. The audit therefore records the window over which the census was taken and the sources used, and it is repeatable on the same rule at a later date. A later census returning the opposite result is a finding about the industry and is reported as such, not as a refutation of this proposition

as originally scored, and the earlier verdict stands against its own window. INSTRUMENT: DESIGN DRAFT, BLOCKED ON AN UNWRITTEN MAPPING FROM THE AUDIT'S CLASS DISTRIBUTION TO THIS PROPOSITION'S THREE VERDICTS, the corrector-class audit.

SUCCESSOR QUALIFICATION. The preceding instrument-status sentence is retained from the version 1.102 snapshot published on 13 September 2026. DATED STATUS, 3 OCTOBER 2026 AT 00:09:45 UTC. Unit 06, the corrector-class audit, was submitted on 1 October 2026 at 17:13:36 UTC under OSF identifier rv2b5. At the stated observation time OSF recorded public=false and pending_registration_approval=true: SUBMITTED, PENDING APPROVAL, not approved and public. All 19 archived files and all 23 text fields matched the deposited version; the six attachment fields also matched by file identity and hash. These are registration and text-conformance facts, not a study run, a proposition verdict or a certification of what the author knew. No study outcome is inferred. The archived-file and text matches do not by themselves establish that the instrument satisfies every condition of P7. Its registered endpoint, mappings and admissibility conditions govern that separate question.

SUCCESSOR QUALIFICATION. DATED STATUS, 11 OCTOBER 2026 AT 19:01:59 UTC. Unit 06, the corrector-class audit, is registered under OSF identifier rv2b5 (DOI 10.17605/OSF.IO/RV2B5): at the stated observation time OSF recorded it public and no longer pending approval, and its public results record it as shown approved and public at 05:10:36 UTC on 4 October 2026. Its registered run took place on 6 October 2026. Run A completed 303 calls; Run B returned HTTP status 402 at its first call, which the registered failure policy makes final, so no row was counted. The study is reported as administratively incomplete under its registration, with no inferential verdict. The results are public at https://osf.io/4bphg/ (DOI 10.17605/OSF.IO/4BPHG), and the registration rv2b5 links to them as its data resource. These are registration and run facts as OSF's records and the unit's public results state them; no proposition verdict is inferred from them. Of P7, the run's frozen analysis printed under-powered: no verdict on P7 stands from the run, and no rung is reached.

THE AUDIT'S OWN PRIMARY IS NOT THIS PROPOSITION'S DICHOTOMY, AND THE TWO CAN DISAGREE IN ONE RUN. This proposition sets the class whose strength is fixed by construction against the class that can scale with the capability it corrects. The audit's own primary hypothesis sets the union of those two against a third class that accumulates its own strength, so a population drawn wholly from the scaling class would satisfy that hypothesis while refuting this proposition. No rule maps the audit's class distribution onto the three verdicts registered above. That mapping is written down before any outcome is retrieved, and until it is, this proposition is NOT EVALUABLE on that instrument rather than scored on a distribution answering a different question.

AND FOUR THINGS THIS ENTRY CLAIMS ARE REQUIREMENTS ON THAT UNIT RATHER THAN PROPERTIES IT ALREADY HAS: the second, deployment-weighted count; the explicit mixed or ambiguous cell, counted and reported rather than allocated; the tie-break rule registered in advance; and the coding of the independence axis beside the scaling axis. They are registered here as conditions the deciding unit meets before its result is admitted against this proposition, and an audit meeting fewer of them decides less. CONFIDENCE: moderate.

P8 · MEASURED EXPONENTS EXCEED THE NULL. Across the admissible independent systems, the pooled growth exponent exceeds the one-half null. The direction is part of the claim: an interval lying wholly below the lower edge of the registered equivalence region refutes this proposition rather than supporting it, which an earlier two-sided wording did not say and the decision rule below has always assumed. The edge and not the null is the boundary here exactly as it is in the rule beneath, and no shorter wording anywhere in this document overrides the four regions that rule fixes. Whether the exponent also exceeds one is reported separately as the CLEARS UNITY verdict.

REFUTED BY: a pooled interval on the growth exponent lying wholly below the LOWER EDGE of the registered equivalence region around one half.

AN EARLIER WORDING SAID ENTIRELY BELOW ONE HALF, AND IT CONTRADICTED THE FOUR-OUTCOME RULE IT SITS ABOVE. An interval from thirty-five hundredths to forty-five hundredths lies entirely below one half while straddling the region's lower edge, so the short form returned REFUTED where the rule returns INCONCLUSIVE and one interval carried two permissible verdicts. The four-outcome rule governs, and equivalence to the null is its own separate verdict for an interval lying wholly inside the region rather than a second route to refutation. An interval that merely contains one half refutes nothing and is reported as INCONCLUSIVE. An earlier wording of this line said the opposite and it is corrected below rather than quietly dropped.

FOUR OUTCOMES, EXHAUSTIVE, WITH THE SCORING VERDICT NAMED BESIDE EACH. The pooled rule is registered here because many per-system intervals feeding one verdict is the pattern this programme has learned to distrust: across the admissible independent systems the exponents are pooled by inverse-variance weighting, and then ABOVE THE NULL, scored SUPPORTED, when the pooled 95 per cent interval lies entirely ABOVE THE EQUIVALENCE REGION, that is wholly above 0.60 on the exponent scale; BELOW THE NULL, scored REFUTED, when it lies entirely below that region, wholly below 0.40; EQUIVALENT TO THE NULL, also scored REFUTED, when the interval lies entirely inside the equivalence region around one half fixed by the equivalence condition in the measurement field, at plus or minus 0.10 on the exponent scale, that is within 0.40 to 0.60; and INCONCLUSIVE in every other case, including an interval that merely contains one half and one that straddles a boundary of the region.

THE BOUNDARIES ARE THE REGION'S AND NOT THE NULL'S, corrected here because an earlier wording put the two clearing verdicts at one half while the equivalence verdict ran to 0.60, so an interval from 0.51 to 0.55 satisfied a support rule and a refutation rule at once and the sentence claiming exhaustiveness was describing a partition the rule did not implement. The four are exhaustive by construction: an interval either clears the region on one side, sits inside it, or does neither.

WHAT ONE HALF IS AND IS NOT, because the number does two jobs in this document and only one of them is this proposition's. Here it is the zero-parameter benchmark that independent-improvement accumulation returns, and this proposition asks only whether measured exponents depart from it. It is not the empirical baseline for the framework's recursion claim: that baseline is the compute-matched breadth arm registered under P1, because breadth also produces smooth

power-shaped gains and a departure from one half by itself does not distinguish depth from sampling. A supported P8 beside a failed P1 RECURSION verdict is reported in exactly those words.

AN INTERVAL CONTAINING THE NULL IS NOT EVIDENCE FOR THE NULL, and an earlier wording of this proposition treated it as though it were. That wording marked the proposition refuted whenever the pooled interval included one half, which confuses a failure to demonstrate a departure with a demonstration of its absence. Refutation now requires either an interval wholly below the lower edge of the equivalence region or an interval tight enough to sit inside it, and the equivalence region is fixed before any result with the instrument required to show it could have detected a departure of that size. How many individual systems returned an interval clear of one half is printed alongside the pooled verdict and is never offered in place of it. The eight-system floor governs support as well as refutation: below eight admissible systems there is no verdict in either direction, including a pooled interval that would otherwise support.

THE FLOOR HERE IS COUNTED IN ADMISSIBLE SYSTEMS AND THE FLOOR P1 REGISTERS IS COUNTED IN ESTIMABLE CELLS, which are not the same object, and an earlier wording in that entry said the two were one count. Which noun governs both entries is the author's to fix and is recorded in the deciding instrument before any outcome is read. Until it is recorded, the interim rule registered with the admissibility conditions binds in the banking direction only.

THE DIRECTION IS PART OF THE VERDICT, and the downward case is named here so that it cannot be absorbed as support: a pooled interval lying entirely below the lower edge of the registered equivalence region, that is wholly below 0.40, is reported as BELOW THE NULL, and it is a refutation rather than a departure to be counted in the framework's favour. It would mean recursion converting effort into capability less efficiently than drawing independent samples, which is a worse outcome for this programme than sitting on the null. An earlier wording of this proposition asked only whether the interval excluded one half, and would have reported that result as a win; the wording is corrected here rather than left to be discovered by whoever the outcome happened to suit.

THE SECOND THRESHOLD, BECAUSE A DEPARTURE ALONE IS NOT WHAT THE FRAMEWORK NEEDS: an upward departure that does not reach one satisfies the conversion claim nowhere, since that claim is super-linear, and the programme's own published position states the live empirical question as whether the exponent can reliably exceed one. A second verdict is therefore registered and reported beside the first: CLEARS UNITY, on whether the pooled interval lies entirely above one. A supported first verdict beside a failed CLEARS UNITY is reported in exactly those words, and never as support for the framework's conversion claim. On the sheet the mark follows the first verdict and the CLEARS UNITY result is printed beside the mark, so a supported first verdict beside a failed CLEARS UNITY is marked supported with those words beside it.

THE CAPABILITY SCALE CONDITION IS APPLIED TO EVERY EXPONENT POOLED HERE, and bears hardest on this second verdict because a capped scale removes the exponents that would clear one. The 22 August 2026 ladder-scale correction governs every unit offered here, on P1's terms.

THE ALLOCATION TITRATION CONTRIBUTES EXPONENTS HERE ON THE SAME TERMS AS UNDER P1. Its sustained growth exponents are admissible among the systems pooled here wherever

they meet the conditions registered above, and none is evidence about a correction-limited boundary, as P2's entry sets out.

FALLS WITH IT: the necessity of the compounding reading. The framework's arithmetic survives but becomes unnecessary, which is a real cost and is stated as one. INSTRUMENT: DESIGN DRAFT, the multi-system exponent estimation. CONFIDENCE: moderate, and lower than the programme's public materials have implied, given that a robust summary of the programme's five earlier runs sits close to the null with an interval that reaches below zero, and that four of those five nonetheless put the sequential arm above the matched parallel arm. Both halves of that sentence are the reason the confidence is neither high nor low.

P9 · THE CROSS-DOMAIN FORM. Where an effective dimension can be assigned to a domain, the growth exponent follows the dimensional relation with the dimension supplied by that domain's equation of state.

WHAT IS ESTABLISHED OUTSIDE THIS PROGRAMME AND WHAT IS REGISTERED HERE, stated before any measurement because the opposite order would be an overclaim. The dimensional form itself is established outside this programme: the relation between an effective dimension and a scaling exponent in space-filling transport networks is the subject of a substantial literature in biological allometry and network scaling, and this registration claims no priority over it. Two things are registered here instead. The first is the assignment: that a domain's effective dimension can be read from how that domain composes its inputs and outputs, rather than from its geometry alone, and that the exponent follows from the dimension so assigned. The second is that the assignment is made before the domain's curve is retrieved. A measurement that agrees with the established form while the dimension was chosen after the fact supports nothing here, and is reported as agreement with the established form rather than as support for this proposition.

THE PROGRAMME'S OWN DERIVATION IS DATED AND IS RECORDED HERE AS PROVENANCE, not as a priority claim over the form conceded above. This programme derives the same dimensional exponent from three conditions, multiplicative composition, d-dimensional space-filling, and a conservation constraint on resource flow, and its reading is that the independently published derivations agree because any derivation meeting those three conditions returns that exponent whatever physical mechanism instantiates them. That reading was put in writing on 24 March 2026, with both of the programme's papers attached, to Professor Lloyd Demetrius of Harvard and Professor Geoffrey West of the Santa Fe Institute, both of them authors of derivations named against this proposition in the references, and no qualified refutation of it is on the record. What the record supports is a send on that date and nothing beyond it: silence is consistent with a letter unread, it is never reported here as agreement, and none of it is evidence for or against this proposition. The concession stands, the relation between dimension and exponent remains the literature's, and what is registered here is still the assignment and the fact that it precedes the curve.

REFUTED BY: measured exponents disagreeing with the dimensional prediction across two or more independent domains.

THE RULE A SCORER APPLIES IS THE INSTRUMENT'S, AND IT IS STATED HERE SO THAT THE TWO CANNOT DIVERGE. The units that would decide this proposition run equivalence tests against a value frozen in advance, with an explicit exclusion of the rival that predicts one exponent for every system regardless of dimension, and with a power gate refusing a discriminating verdict where the observed spread makes the comparison underpowered. Both are carried here: no verdict is returned on this proposition unless that rival is excluded on the same cells, and an underpowered comparison is reported as inconclusive rather than as agreement.

THE TOLERANCE AND THE DEFINITION OF AN INDEPENDENT DOMAIN ARE NEITHER OF THEM SETTLED HERE. Neither is registered in this document. Both are the author's, both are recorded in the deciding instrument before any outcome is read, and until they are this proposition is NOT EVALUABLE, which is the status this entry already carries for the unwritten assignment rule.

FALLS WITH IT: the cross-domain derivation only. The empirical programme is untouched. Nor does this proposition fall with P1, which is the reason the dependency map above gives and this entry now honours: the exponents this proposition concerns are measured in other domains and exist whether or not software recursion follows a power law, so what a failed P1 costs this proposition is its reading as an instance of the same conversion law and nothing more.

NOT EVALUABLE until the rule that assigns an effective dimension to a domain exists as a written procedure, for the same reason as the composition proposition below, which is blocked at the next stage of the same requirement: there the rule is written and its transmission to a third party is unmeasured, while here the rule is not written at all. The assignment is characterised in this document and it is not specified: no rule is written such that a third party applying it to a domain description returns the same dimension, so any assignment made today would be the author's judgement applied case by case, which is the retrospective classification this proposition would have to exclude in order to carry weight. Consistency requires the same verdict here as there, NOT EVALUABLE on both, though the block sits at a different stage of the same requirement, and it lowers what this set can currently claim. Earlier versions of this proposition recorded no instrument in existence and none in draft, and that was wrong against this programme's own records: two drafted units test the dimensional relation. One fits the framework's dimensional exponent against the competing quarter-power law on published biological data, with two separable predictions at effective dimensions of one and two. The other tests, per dimensional class, whether the fitted exponent is equivalent to a value frozen in advance, against a rival that predicts one exponent for every system regardless of dimension. What neither supplies is the part of this proposition that specifies the dimension from the domain's equation of state rather than from its geometry, so this proposition is decided by those two only in the part they cover, and the remainder waits on a design that derives the dimension the way the statement requires. The author regards this as the most speculative line in the framework and its failure as the cheapest. INSTRUMENT: DESIGN DRAFT, BLOCKED ON A FROZEN DIMENSION-ASSIGNMENT RULE, on a correction made here. The surrounding design is written; the rule that assigns an effective dimension to a domain is not, and the status names that rather than implying a runnable draft. CONFIDENCE: low.

P10 · SCORER-FAMILY DEPENDENCE. A material fraction of the results this programme has scored unblinded within a single model family fails to survive re-scoring by a different family, where material is the ten per cent fixed below. An earlier wording said only that not every result survives, which a single failure would satisfy while the rule below scores something stronger; the statement and the rule now carry the same claim. The eligible corpus and the unit counted as one result are fixed by the deciding unit's own registration before any re-scoring is run.

REFUTED BY: a survival proportion whose interval excludes a material loss, where material is fixed now at ten per cent, which is a declared convention fixed before any result rather than a quantity this framework derives, so a fraction landing just either side of it is reported as a result at the margin: P10 is REFUTED when the 95 per cent interval on the fraction of earlier results that do not survive cross-family re-scoring lies entirely below ten per cent, SUPPORTED when it lies entirely above ten per cent, and INCONCLUSIVE otherwise. The re-scoring study's own registration carries the same threshold. A refutation here would mean this programme's blinding result does not reach its own earlier work, which is the one direction in which that result is convenient for its author.

A FOURTH OUTCOME IS IMPORTED FROM THE DECIDING UNIT, IN ITS OWN TERMS, because that unit registers four where this entry registered three. Below a floor of decidable results it refuses to score at all and returns NOT EVALUABLE, which had no home here, so a scorer reading this document alone would have reported inconclusive for a case the instrument declines to score. NOT EVALUABLE is registered here as the fourth outcome, is marked untested on the sheet with the decidable count printed beside it, and is never reported as inconclusive.

AND THE FROZEN ROSTER'S DECIDABLE COUNT IS PRINTED BESIDE THE VERDICT, because the deciding unit records that declaring a corpus not implicated needs about thirty-six decidable results with none lost, while a small corpus with few losses returns inconclusive whatever the truth. Without that count nobody can tell whether the refutation this entry registers against its author's interest is arithmetically reachable at all, and the claim that it runs against his interest rests on its being reachable.

WHAT A SUPPORTED RESULT HERE WOULD LICENSE, in the restrictive form the licence rule above fixes: it licenses the statement that a material fraction of the results this programme has scored unblinded within a single model family fails to survive re-scoring by a different family, at the fraction this entry fixes as material, and it does not license any statement about any other corpus, about which of the two scorings is correct, or about the scorer itself.

FALLS WITH IT: nothing in the framework. What falls is the evidential standing of the earlier results themselves, which is why this prediction is registered against the author's interest. INSTRUMENT: DESIGN DRAFT, the cross-family re-scoring unit. The cheapest of the three cases named above as the ones that would decide most, because it generates no new model responses. CONFIDENCE: moderate that more than the registered material fraction of the frozen eligible corpus, ten per cent, changes verdict under re-scoring by a different family.

P11 · THE RESIDUAL-DECAY AND CORRECTION-CAPACITY EXPONENTS ARE NOT PRACTICALLY INTERCHANGEABLE. The residual-decay exponent and the correction-capacity exponent are not practically interchangeable across the systems and conditions specified in advance, under the registered equivalence margin, and the programme's working practice of treating them as one is an assumption rather than a fact.

THE CLAIM IS INTERCHANGEABILITY AND NOT IDENTITY, because the second cannot be refuted by a measurement. An earlier wording said the two are distinct quantities and offered numerical agreement as its refuter, which does not follow, and the title of this entry carried that withdrawn word until this release: agreeing exponents would leave two constructs agreeing, not one construct. What is registered here is the weaker and testable claim, which practical equivalence under the registered margin does refute. Agreement never establishes that the constructs are the same, and never licenses substituting one for the other inside a derivation.

NARROWED TO THE PAIR ITS INSTRUMENT ACTUALLY COMPARES. An earlier wording named three exponents while the refuter and the named instruments compare only two of them, so the proposition claimed more than it could decide, and the narrowing left the statement itself split across the correction. Both are repaired here. The capability-to-depth conversion exponent remains a third and separate quantity in the notation extract above, and no claim is registered here about its relation to either of the other two.

REFUTED BY: measurements of the first two, taken on instruments that share no fitted quantity, agreeing within a registered tolerance across model families, which would license the identification the programme currently only assumes.

THE TOLERANCE IS NOT YET A NUMBER, and this proposition is undecidable in either direction until it is. None is registered for it in this document or on either side of the deciding design, so the warning the equivalence condition gives, that an agreement claim whose tolerance is chosen after the data is not a claim at all, applies to this proposition in full. The tolerance is the author's to fix, it is recorded in the deciding instrument before any outcome is read, and until then this proposition is NOT EVALUABLE.

SHARING NO FITTED QUANTITY IS NOT ENOUGH, AND THE CLAUSE IS WIDENED IN THE STRICT DIRECTION. Two instruments that share no fitted quantity may still share their sample, their capability ladder, their fault set, their seed stream and their scorer, and under that sharing a common measurement error pushes the pair towards the agreement that is the scored outcome here, which is the error-on-both-axes failure the measurement field registers. Where both estimates are taken under shared nuisance, the sharing is named beside the verdict and the agreement is reported with that limitation attached, never as a clean refutation.

WHAT A REFUTATION HERE WOULD AND WOULD NOT LICENSE, registered before any result because the temptation runs one way. Numerical agreement between two exponents is not identity of the quantities that produced them. Two genuinely distinct constructs can carry the same exponent over a tested range, and two measurements of one construct can differ through noise, bias or a shift of domain. A refutation of this proposition therefore licenses the working practice of treating the

two as one for estimation, and it does not license any statement that they are the same quantity, nor any derivation that substitutes one for the other. Before either exponent may stand in for the other inside the ceiling relation or any other derived expression, an intervention is required that is designed to move one of them while leaving the other where it was, and the identification is licensed only if that intervention fails to separate them. Agreement without such an intervention is recorded as consistent with identification and never as identification.

FALLS WITH IT: nothing. Refutation here would strengthen the framework by converting an assumption into a result, which is why it is registered. INSTRUMENT: DESIGN DRAFT, on both sides of the comparison. CONFIDENCE: not stated. The author has no view worth recording.

P12 · BUILD ORDER MOVES THE CORRECTION-LEVERAGE EXPONENT. The order in which a system's components are built changes its correction-leverage exponent by at least the registered minimum effect. An earlier heading read that build order moves the class, which is the reading this proposition's own body withdraws, and a heading is the formulation most likely to be quoted away from the body that qualifies it.

REFUTED BY: a preregistered manipulation of build order producing no change in the correction-leverage exponent beyond the tolerance registered with that instrument, where the instrument has shown from its own sensitivity that a change of that size would have been detected. A manipulation too weak to move anything is not a refutation of this prediction; it is a failed instrument, and condition D above says so in advance.

THE MANIPULATION IS NAMED SO THAT A NULL CANNOT BE EXCUSED BY IT: the registered maximal contrast is a corrector formed before the capability it corrects against a corrector added after it, at matched content, and a null on that contrast refutes this proposition rather than counting as a pair of build orders that were not different enough. Without a defined manipulation that excuse is always available and the proposition is unfalsifiable in both directions, which is the state this clause ends.

ONE ENDPOINT, ON ONE SCALE, and the earlier version of this paragraph had the wrong scale. The endpoint is the correction-leverage exponent, the quantity the third law's ceiling is written in, and the claim is that build order moves it by at least the registered minimum effect, which is the equivalence margin this document registers on the exponent scale and is therefore a declared convention rather than a derived quantity, held in common with the other exponent comparisons so that build order is not judged on an easier scale than they are. It is a continuous claim and it is scored as one. Any reading of this proposition as a crossing between corrector classes is withdrawn: a class boundary on a continuous exponent would need a conversion rule frozen in advance, no such rule is registered, and calling a continuous movement a class change without one is how a modest effect gets reported as a qualitative result. The mechanism the framework proposes for the effect is a change of class; the quantity registered and scored is the exponent, and no class language appears in the title, the refuter or the scoring.

THE MINIMUM EFFECT OF INTEREST, fixed now because a tolerance set later by the instrument that uses it would be circular: as written the refuter turned on a tolerance registered with an

instrument that does not exist, so whoever built it could choose a tolerance that decided the answer. An earlier wording fixed that minimum at 0.05 on the leverage-fraction scale and called it the programme's own registered default. The number is real and the scale was wrong. That default is registered by the substrate-ceiling unit as a margin on the self-referential coupling, the leverage fraction of the first law's growth equation, which is a different quantity from the correction-leverage exponent this proposition moves. Importing a margin across two quantities that share the word leverage is the exact substitution this document forbids elsewhere and had committed here. The minimum effect of interest is therefore fixed at 0.10 on the correction-leverage exponent scale, which is the equivalence margin this document registers for every exponent comparison it decides, applied here rather than borrowed from a neighbouring quantity. A future instrument may register a smaller minimum and may not register a larger one, and a manipulation moving the correction-leverage exponent by less than this is reported as no effect of interest rather than as no effect.

FALLS WITH IT: the programme's argument that build order is an intervention point rather than an accident of history, and this proposition's share of the protocol's engineering case, which it carries jointly with P13, P15 and P21.

SUCCESSOR QUALIFICATION. Under successor version 1.103 this proposition is also read as the build-order limb of Law V, with a direction it does not itself carry: the limb is supported where P12's interval lies wholly at or beyond the minimum effect on the side of the corrector formed first, and refuted where it lies wholly at or beyond the minimum effect on the side of the corrector added after, or where the refuter above fires; a refuted limb refutes Law V. A P12 supported in the direction of the corrector added after therefore refutes Law V while P12 keeps the mark it earned: it establishes that build order is an intervention point, it goes against the method's direction, which is correction formed first, and it is never reported as support for the method (line 142). P12's wording, refuter, minimum effect, instrument line and mark are unchanged.

WHAT THIS PROPOSITION IS IN THE FRAMEWORK'S OWN QUANTITIES. The general boundary is set by four quantities, and this proposition is a claim about the same one P22 concerns, the correction elasticity, reached by a different route: P22 says that elasticity differs with the type of fault corrected, and this says it moves with the order in which the corrector was built. Where it moves upward the boundary rises with it. The two are severable and neither carries the other. INSTRUMENT: NONE, and the reason is narrower than it looks. Registered without one for the same reason P5 was, before the design that entry specified was drafted. Several drafted units manipulate placement or build order, on gate position, on whether a value specification appears first or last, and on formation against a later rule, but none of them measures the correction-leverage exponent, which is this proposition's endpoint. This proposition is therefore one endpoint away from an instrument rather than lacking a design, and that is recorded here so a reader who finds those units does not conclude the statement was careless.

WHAT A SUPPORTED RESULT HERE WOULD LICENSE, and what it would not, in the restrictive form the licence rule above fixes: it licenses the statement that the order in which a system's components are built changes its correction-leverage exponent by at least the registered minimum effect, on the registered maximal contrast, and it does not license any statement about where the boundary sits, about a system built in that order being safer, or about orders outside that contrast.

WHAT WOULD DECIDE IT, SPECIFIED RATHER THAN LEFT AS AN ABSENCE, on the model of the neighbouring proposition that specifies its own missing extension: a typed extension of the correction-exponent estimator that fits corrective strength against overseen capability and reads the correction-leverage exponent across a registered maximal build-order contrast, with both orders estimated under one instrument so that their difference carries a joint interval. That build-order extension is not built, does not exist anywhere in this programme's own work, and will be registered as its own unit or as a recorded amendment to the estimator's own registration before any outcome is examined. Specifying what is missing is what caused P5's instrument to exist, and an absence that names what would fill it is worth more than an absence that does not. CONFIDENCE: low.

P13 · SAME-FAMILY AUTOMATED CORRECTION UNDERPERFORMS CROSS-FAMILY AT MATCHED RESOURCES, AND THROUGH THE MEASURED ERROR CORRELATION.
Where an automated alignment researcher post-trains a target model against alignment benchmarks, a researcher from a different model family than the target reduces held-out misalignment more than a researcher from the same family, at matched data and compute, and the panels differ in measured pairwise error correlation in the direction that would explain it. This is one conjunctive proposition and not two, stated so in its first sentence because the rule below scores it that way: an advantage without the corresponding difference in correlation is reported as unexplained by this framework and never as support. This is the corrector-class contrast, stated for the apparatus of Chen, Wen and Kirchner (arXiv:2608.28945), whose main runs are cross-family and whose same-family run puts a Claude Sonnet 5 researcher on an early Claude Opus 4.8 checkpoint. That run's winning method scores 65 per cent on the paper's own audit against 72 per cent for the released model after its full production alignment training, using about 2,400 training examples. An earlier wording of this entry described those two figures as fractions of a safety gap closed. They are not: they are scores on that audit, and the correction is recorded here rather than left in a document whose credibility rests on a reader being able to check it. That pair is not evidence either way for this proposition, because the two runs are not matched on data or compute, and the matched comparison is the experiment this prediction requests.

REFUTED BY: the interval on the cross-family minus same-family difference in held-out misalignment reduction sitting wholly at or below zero, with data and compute matched and with scoring blind to condition.

TWO VERDICTS, BECAUSE ONE RESULT WAS FALLING BETWEEN THEM. The performance axis is the cross-family minus same-family difference in held-out misalignment reduction, and it is scored supported, refuted or inconclusive by the interval rule above: that axis is this proposition as worded. The mechanism axis is whether the registered pairwise-correlation contrast between the panels moves in the direction that would explain the difference, and it is reported supported, unsupported, inconclusive or NOT EVALUABLE beside it, the last being the value it carries for as long as no unit measures that contrast. An advantage appearing beside a measured correlation contrast that does not move in the explaining direction is a supported performance axis with an unsupported mechanism axis, reported in those words and never as support for the framework's explanation; an earlier wording left that cell as neither support nor refutation, which is not a verdict. Where no unit

has measured that contrast at all, which is the position today, the mechanism axis is NOT EVALUABLE rather than unsupported, and the two are never reported in the same word.

FALLS WITH IT: the different-substrate engineering recommendation, and this proposition's share of the protocol's engineering case, which it carries jointly with P12, P15 and P21. No law falls; the independence premise and the ceiling relation are untouched.

THE MECHANISM IS THE ERROR CORRELATION AND NOT THE FAMILY LABEL, and it is measured rather than assumed. Family is a proxy for what this framework says matters, which is how far two correctors fail together. The equicorrelation model registered in the alternative baselines is written in that correlation, and the published work named against the independence premise measures it directly across large model panels. This proposition is registered as holding through the correlation or not at all, and the mechanism axis has no instrument on this apparatus. No unit in this programme measures the common pairwise error correlation of a same-family panel and of a different-family panel on the systems this proposition is stated for, so that axis returns NOT EVALUABLE and is reported in that word rather than inferred from the performance axis.

NEITHER AXIS IS DECIDED BY ANY UNIT OF THIS PROGRAMME AS THIS PROPOSITION IS WORDED, AND THE INSTRUMENT LINE IS CORRECTED HERE TO SAY SO. The unit previously named against this proposition estimates a target-balanced difference in a residual-decay exponent between corrector classes, which is not the held-out misalignment reduction this proposition is worded in; it returns an association where this proposition asks for a matched comparison; and its own text disclaims establishing an error-correlation mechanism and routes mechanism elsewhere. An earlier wording said that design applies directly, which claimed more of it than it does, and a later one kept it as this proposition's instrument, which the status rule fixed above does not permit: a proposition's status is the status of the instrument that would decide it and never of a neighbouring unit that would inform it. Because this proposition is conjunctive, it cannot be marked supported on the scoring sheet while either axis is without an instrument. It is marked untested there, with the missing matched comparison and the missing correlation contrast printed beside the mark, and refutation on the performance axis remains available exactly as the interval rule above states it. A supported performance axis beside a mechanism axis returning NOT EVALUABLE is the best cell attainable once both units exist, and it is not a cell attainable today. INSTRUMENT: NONE, corrected here from a value the neighbouring unit could not carry.

WHAT WOULD DECIDE IT, SPECIFIED SO THAT AN ABSENCE BECOMES A REQUIREMENT: on the performance axis, a matched-resource corrector-class unit measuring held-out misalignment reduction at matched data and compute on the systems this proposition is stated for, with scoring blind to condition; and on the mechanism axis, a paired correlation unit measuring the common pairwise error correlation of both panels against the same target on the same held-out misalignment measure, so that the difference in correlation and the difference in reduction are read on one instrument. The published apparatus this proposition responds to runs the right kind of experiment, and its own same-family and cross-family runs are not matched on data or compute, which is why the matched comparison is the experiment this prediction requests rather than one it can be scored on.

AND THE CORRECTOR-CAPABILITY COVARIATE IS CARRIED HERE AS WELL AS THERE, since that unit registers corrector capability as a diagnostic: a cross-family advantage explicable by the corrector simply being the stronger model is reported as unexplained by this framework, in the same words the correlation rule above uses, and never as support for it. CONFIDENCE: moderate.

P14 · GAINS SHRINK UNDER BLINDED CROSS-FAMILY SCORING. Where an automated researcher's method is re-scored by a scorer that cannot see the condition and comes from a different model family than either the researcher or its target, the alignment gain it returns is smaller than the gain the unblinded same-family scoring reported.

REFUTED BY: the interval on the blinded minus unblinded gain sitting wholly at or above zero across the same items.

THE MARGIN THIS REFUTER IS READ AGAINST IS NOT YET A NUMBER, and it is stated here rather than left to the equivalence condition alone. That condition names margins for other propositions and none for this one. The margin is the author's to fix, it is recorded in the deciding instrument before any outcome is read, and until it is, this proposition is NOT EVALUABLE on that ground as well as on the released per-item outputs its instrument line names below.

SUCCESSOR QUALIFICATION. INSTRUMENT: NONE under successor version 1.103, corrected by this qualification from the value in the instrument line below, which is retained as registered. The design that line names runs a redaction factorial on this programme's own frozen scenario bank, as this entry records below, and does not estimate the difference in alignment gain between blinded cross-family and unblinded same-family scoring of the same items, which is what the refuter above reads. A proposition's status is the status of the instrument that would decide it and never of a neighbouring unit that would inform it (line 406), so this successor counts P14 with P12, P13, P15 and P22 among the propositions with no deciding instrument, and the scoring sheet prints NONE in its instrument column. The required design scores the automated researcher's method under matched blinded cross-family and unblinded same-family regimes on the same items, estimates the difference in alignment gain, and registers its operating characteristic and the author's margin before any run. The claim, the refuter, the instrument line and the confidence are unchanged.

FALLS WITH IT: nothing in the framework. What is lost is the evidential standing of alignment scores taken unblinded within one model family, which is this programme's own published blinding result turned on somebody else's. INSTRUMENT: DESIGN DRAFT, the blinding-evaluation design.

THE REFUTER DEPENDS ON ANOTHER TEAM'S PER-ITEM OUTPUTS, AND THE DEPENDENCY IS REGISTERED ON ITS FACE. The comparison above is stated across the same items, and the blinding-evaluation design runs a redaction factorial on this programme's own frozen scenario bank with its own engines rather than on those items. Where the per-item outputs of the published apparatus are not released, this proposition is NOT EVALUABLE rather than scored on a substitute corpus, in the same terms this document uses wherever a proposition waits on a quantity nobody has measured.

WHAT A SUPPORTED RESULT HERE WOULD LICENSE, stated in the restrictive form the licence rule above fixes: it licenses the statement that where an automated researcher's method is re-scored by a scorer that cannot see the condition and comes from a different model family than either the researcher or its target, the alignment gain it returns is smaller than the gain the unblinded same-family scoring reported, and it does not license any statement about the method's true effect or about the blinded scorer being right.

AND THE ENDPOINT IS A DIFFERENCE OF GAINS ON A BOUNDED RUBRIC, which a uniformly stricter blinded scorer can shrink by compressing the top of the scale, with no truth about blinding involved. The deciding unit therefore reports the gain on a rank scale, or under a registered anchor-invariance check, beside the raw scale, and a difference surviving on the raw scale alone is reported as scale-dependent. CONFIDENCE: moderate.

P15 · EXTERNALLY INSTALLED GAINS DECAY UNDER SUBSEQUENT CAPABILITY TRAINING. Gains installed by external post-training erode under later capability-directed training that embeds no correction of its own, and more than half of what was installed is lost once the capability-training budget matches the alignment budget that installed it. The one-half threshold is a declared convention fixed here before any result rather than a derived quantity, in the same way P10 fixes ten per cent. The authors of arXiv:2608.28945 state this as their own untested question, quoted verbatim from the third version of that paper: "we do not test whether gains persist after extensive reinforcement learning on other tasks." An earlier wording of this entry paraphrased that sentence while presenting it inside quotation marks. The paraphrase was faithful in sense and was not the authors' words, and in a document that asks a reader to check every citation it is the kind of error that costs more than it saves.

REFUTED BY: an interval on the retained fraction of the installed gain, after matched capability training without embedded correction, lying entirely above one half.

FALLS WITH IT: this programme's case that correction has to sit inside the system rather than be applied to it from outside, and this proposition's share of the protocol's engineering case, which it carries jointly with P12, P13 and P21.

WHAT A SUPPORTED RESULT HERE DOES NOT ESTABLISH, said because the converse reading is the easy one: support shows that externally installed gains lack the registered persistence under the matched subsequent-training regime, and it does not show that embedded or in-loop correction is superior, sufficient or safe. The comparison between placements is P21's, not this proposition's. The laws are untouched. INSTRUMENT: NONE, corrected here from a value this document's own definition made false. The designs previously named against this proposition cross correction regimes with recursive depth on frozen provider endpoints and update no weights anywhere, while this proposition is about the persistence of a gain installed by post-training, so nothing in this programme measures its endpoint and the honest status is the one reserved for that state.

SUCCESSOR QUALIFICATION. Here "The laws are untouched" means Laws I, II and III, named when these words were registered. Under successor version 1.103 this proposition is the retention limb of Law IV, so a refuted P15 refutes Law IV. It does not strike Law V, which only a refuted verdict on Law

IV's comparison limb can strike, and it scores neither P12 nor P21. The registered comparison between placements remains P21's; Law IV's comparison limb, which has no registered proposition yet, compares the fractions two installations retain after the same training.

WHAT WOULD DECIDE IT, SPECIFIED SO THAT AN ABSENCE BECOMES A REQUIREMENT: a two-arm persistence unit in which a gain is installed on one arm by external post-training and a matched gain is installed on the other with correction embedded in the training objective, both arms then undergo the same capability-directed training, and the retained fraction is read on both, so that the contrast and not a single arm licenses the reading about placement. That unit fixes the budget in which the capability-training budget matches the alignment budget that installed the gain, since the one-half threshold is otherwise decided by an unstated choice of scale rather than by nature. The budget is the author's to fix, it is recorded in the deciding instrument before any outcome is read, and until he fixes it the threshold is NOT EVALUABLE on that scale. The unit also carries an attestation that the capability-training reward includes no alignment-graded component, because the statement above specifies training that embeds no correction of its own and nothing at present verifies it. CONFIDENCE: moderate.

P16 · THE PROHIBITION ITSELF. A system does not hold both of these at once: a growth exponent above the point at which its own measured correction margin reaches zero, and a correction margin whose lower bound stays above zero, sustained across a window fixed before the run.

WHY THIS IS REGISTERED SEPARATELY, when P3 already carries the conjunction in its refuter: a conjunction that lives only inside a refuter can fail to be refuted, and can never be supported. An excursion above the ceiling that destabilises would have scored as P3 not refuted, which is an absence of evidence against rather than evidence for. This framework's most distinctive claim was therefore registered as a defence. It is stated here as a prediction so that the observation the framework actually expects can be recorded as supporting it.

SUPPORTED BY: an observed run in which the exponent rises above that point and the correction margin then goes negative, in that order, the loss of correctability following the excursion rather than preceding it, replicated at the same settings.

REFUTED BY: a run holding the exponent above that point with the lower bound of the correction margin above zero across the registered window, replicated on fresh data at the same settings.

ONE DECIDING UNIT CARRIES THIS PROPOSITION AND IT RUNS TWO ARMS THAT ARE NOT INTERCHANGEABLE, both of them arms of the unit this entry's instrument line names below, so that this entry names one deciding unit and not several. The primary arm is a titration in which the exponent is driven across a boundary located beforehand from other systems, with dose arms either side of it, a coefficient-only sham arm and an untouched baseline, and with the predicted balance line sealed before the switch. It is the primary because it can refute the boundary's existence and because the experimenter controls the dose and the horizon rather than waiting for a natural trajectory to oblige. The secondary arm is the transport of a boundary forecast, sealed per system from that system's own early checkpoints, onto its later trajectory; it is the bridge to systems nobody

is driving, and it cannot refute an absence. A supported verdict at the level of this proposition requires the titration supported and the transport not refuted. The cell in which the driven arms support the prediction and the transport fails is reported under that description and is never folded into support, because a boundary that appears only when a system is driven towards it is a narrower claim than the one registered here. On the sheet that cell is marked inconclusive, with that description printed beside the mark.

WHY IT DOES NOT WAIT ON THE DERIVED VALUE: this proposition reads each system's own measured margin rather than the value the ceiling relation returns, so it does not inherit the block that holds P3 at NOT EVALUABLE. It carries a block of its own, named in its instrument line below, and no other proposition's instrument reporting discharges it.

THE BOUNDARY MUST BE ESTIMATED ON DATA THIS PROPOSITION IS NOT TESTED ON, or the prediction is definitional rather than prospective. The point at which the correction margin reaches zero is itself estimated. If it is estimated from the same trajectory on which the subsequent loss of correctability is then observed, the proposition reduces to the observation that a curve crosses zero and afterwards lies below it, which is arithmetic rather than a prediction. The boundary is fixed by one of three routes, chosen and recorded before the deciding run: a separate calibration set of systems; earlier runs registered before this proposition; or a registered split in which the boundary is estimated on one part of the data and the prohibition tested on a held-out part. A boundary estimated from the observations on which the prohibition is tested is descriptive, is reported as descriptive, and cannot support this proposition. The framework's most distinctive claim rests here, so this proposition carries the strictest condition in the document rather than the loosest.

DEPENDS ON: P4, and on a local regularity rather than on P1. An earlier version of this entry made it depend on P1, on the reasoning that the exponent this prohibition holds above a boundary has no subject unless capability scales as a power of depth across the tested class. That is the global-form argument the dependency map withdrew, and the map and the scoring sheet both record this proposition as surviving P1's failure. What it needs is weaker and local: that the growth exponent is approximately constant across the window in which each system is estimated, so that a value can be read where the margin reaches zero. On P4, because the correction margin whose zero this proposition locates is P4's construct: if the margin is not the quantity that governs whether a system stays correctable, the boundary is a crossing of a line that means nothing. Both dependencies are stated here rather than left to the map, because this is the proposition the framework's distinctive claim rests on and a reader is entitled to see what it stands on without turning a page.

THE CONTROLS THAT KEEP IT A PREDICTION, REGISTERED WITH THE HELD-OUT RULE ABOVE. Two are required of the deciding unit and are named now. First, cells that do not cross the boundary are run and reported alongside those that do, so that a later loss of correctability is shown to follow the crossing rather than to be the common fate of every long run. Second, the outcome is the time or depth at which the margin turns negative rather than the bare fact that it did, so that a prohibition supported by a system failing eventually is distinguished from one supported by a system failing when the boundary said it would.

FALLS WITH IT: P3, whose ceiling would then bound nothing, and the framework's central prohibition. P1, P2 and P4 survive as measurements. INSTRUMENT: DESIGN DRAFT, BLOCKED ON CANONICAL IDENTITY AND ON DECISIONS THE AUTHOR HAS NOT YET TAKEN, the direct boundary-mapping unit taken jointly with the exponent measurement. CONFIDENCE: moderate to high. The author states his confidence before submission rather than leaving it to be inferred from the framework's enthusiasm.

P17 · PANEL CORRECTOR FAILURES ARE CONDITIONALLY INDEPENDENT ON A FROZEN FAULT SET. Within a corrector panel named in advance, pairwise joint misses on a frozen fault set of stated type are practically equivalent to the conditional-independence baseline under the registered margin.

THE TITLE NOW NAMES THE ESTIMAND THE INSTRUMENT MEASURES, which two earlier titles did not. The framework's premise is about corrections a system accumulates over time; the deciding unit measures coincident misses by several correctors at one moment, which is a cross-sectional dependence and a different estimand. This proposition is now the cross-sectional claim, which the instrument can decide, and it is evidence bearing on the temporal premise rather than a test of it. A registered bridge between the two would be its own unit and does not exist.

THE POPULATION IS NAMED BEFORE THE OVERLAP IS MEASURED, and two earlier wordings of this proposition got that wrong in opposite directions. The first said same-substrate, which names a proxy for the property rather than the property. The second replaced it with the population defined by failure-mode overlap with the generator, which names the property and selects the sample on the very quantity the instrument then measures. Defining a population by an outcome guarantees the outcome, and a proposition supported that way would have been supported by its own sampling frame. The population is therefore a corrector panel named in advance by attributes fixed before any fault is scored: model lineage, substrate, training provenance and the scaling class the corrector taxonomy assigns, all recorded ex ante. Failure-mode overlap is the outcome, not the entry criterion. Substrate enters as an ex ante attribute and as the contrast the deciding instrument uses to attribute any excess dependence, never as the definition of the population. What this proposition measures is how far correctors' blind spots overlap one another on the frozen fault set, and that is the quantity the panel may not be selected on, because defining a population by the outcome guarantees the outcome. What the bound is argued from is a different quantity, how far a corrector's blind spots overlap the generator's, and no instrument registered in this programme measures it: the deciding unit carries no generator arm, both of its panels being correctors reviewing an authored fault set. This matters twice over, because P7's population is the other axis entirely, the corrector's scaling class, and the two were once named as one.

SUCCESSOR QUALIFICATION. WHAT A CORRECTOR IS, READ BY SUCCESSOR VERSION 1.103 BEFORE ANY DATA THAT COULD DECIDE THIS PROPOSITION. For P17 a corrector is one engine under one registered review procedure at one registered reasoning setting, and the four attributes this entry names are recorded for each corrector before any fault is scored. Several correctors drawn from one engine therefore share all four, which places a panel of them in "the population of correctors most likely to share blind spots" (line 1144), where this entry says the premise is cheapest

to attack. The reading is fixed here so that whether such correctors count as several is settled before any joint miss is seen, and a unit deciding P17 states the same definition as its unit of analysis, as condition D requires (line 484). It changes neither the population above nor the refuter.

THE FAULT POPULATION IS AUTHORED AND CALIBRATED FOR DETECTABILITY, AND THAT LIMIT IS REGISTERED HERE RATHER THAN LEFT FOR A READER TO FIND. The population named above is the corrector panel. The faults are a second population and they are not drawn from the world: the deciding unit builds a frozen fault set to a target band of individual catch rates, calibrated on a corrector belonging to neither panel, so that the set is neither too easy nor too hard for correlation to show at all. Some fault population has to be chosen and that choice is not avoidable; the disclosure is. The step from an authored fault set calibrated for detectability to the faults a self-improving system actually generates is registered nowhere in this programme, and no unit bridges the two. A verdict here is therefore reported with that limit attached, in those words: it is evidence about coincident misses on the fault population measured, and not about the fault population a system generates for itself.

WHAT THIS PROPOSITION IS ABOUT AND WHAT IT IS NOT, stated because the instrument measures one of them. The deciding instrument measures coincident misses by several correctors on one frozen fault set: a cross-sectional dependence between correctors. The framework's premise is about corrections a system accumulates over time behaving as though independent. These are not the same estimand, and no result here is offered as a measurement of any correction exponent. What a result here does is parameterise the panel-dependence model and bear on the temporal premise behind P6, in the population where that premise is cheapest to attack, which is the population of correctors most likely to share blind spots. It does not itself establish temporal independence of accumulated corrections: that bridge is untested and would need its own registration. A favourable result here therefore supports one component of the condition set from which one half is derived, and not the value itself.

WHY IT IS REGISTERED: the premise P6 rests on stood undefended, and this document named P6 as the framework's weakest link for that reason, until it recorded that the weak point is the undefended premise rather than the bound resting on it, and named this proposition as the weak point instead. This proposition is the cheapest measurement bearing on that premise, and it is not the premise itself. It does not have to stand undefended. It is measurable, a drafted unit measures it on a frozen fault set committed in advance, and leaving it as an admission understated what could be tested. It is registered in the direction the framework needs, so that a correlated result refutes it.

REFUTED BY: a registered panel drawn from the population above, whose joint failure rate exceeds the independence baseline by the equivalence margin registered with that instrument, which that unit fixes at 0.20 on the log scale of the ratio of observed to predicted joint misses, an inflation of about 22 per cent over independence, on a frozen fault set whose type is stated. The panel and the fault set carry their type, checkable or judged, because P22 predicts the answer differs between them and a panel reported without its type cannot be read against either.

NOT SUPPORTED BY A WEAK PANEL: a panel too small or a fault set too easy to reveal correlation returns INCONCLUSIVE, never support, and the instrument must show from its own sensitivity that

correlation of the registered size would have been detected.

THE VERDICTS HERE MAP ONTO THE DECIDING UNIT'S OWN FOUR LABELS, registered in advance so that the mapping cannot be chosen once the labels have been read. That unit scores each primary pair on the log ratio of the observed joint-miss rate to the rate conditional independence predicts, against the equivalence margin of 0.20, and returns INDEPENDENT where the unadjusted interval lies wholly inside the margin, CORRELATED where the adjusted interval lies wholly above it, ANTI-CORRELATED where the adjusted interval lies wholly below it, on the asymmetry the rule below states, and INSUFFICIENT PRECISION otherwise, in those words. This proposition is SUPPORTED only where every primary pair returns INDEPENDENT, REFUTED where any pair returns CORRELATED, and INCONCLUSIVE where the panel returns INSUFFICIENT PRECISION or ANTI-CORRELATED, the first being what an interval wider than the margin, an interval straddling one of its boundaries, or an undefined interval returns, and the second what a panel systematically diverse in its blind spots returns, and neither of the two is ever read as agreement.

WHERE A PANEL RETURNS MORE THAN ONE OF THOSE LABELS, CORRELATED GOVERNS, so a single correlated pair is a refutation whatever the other pairs return, and INSUFFICIENT PRECISION reaches the sheet only where no pair returns CORRELATED; the strict direction is fixed here so that no refutation is lost to a mixed panel. All four of the unit's labels therefore carry a mark, and every panel the unit can return carries one mark and not two, so the four marks the scoring sheet calls exhaustive hold every outcome this unit can return as a mapping rather than as coverage.

SUCCESSOR QUALIFICATION. WHEN A CORRELATED PAIR IS A REFUTATION, READ BY SUCCESSOR VERSION 1.103 BEFORE ANY DATA THAT COULD DECIDE THIS PROPOSITION. This document's general rule is that "a single cell exceeding a bound is not treated as refutation until it replicates on fresh data at the same settings" (line 516), and the preceding paragraph makes a single correlated pair a refutation; this reading states how the two combine for P17, so that a scorer need not choose between them once labels are in view. A pair that returns CORRELATED on the deciding fault set is a refutation banked under line 516 only when the same pair's 95 per cent interval lies wholly above the registered margin on a second fault set of the same construction, sealed before any corrector reviewed any item and reviewed at the same settings in the same window; a correlated pair that does not replicate is reported as inconclusive, with its first label printed beside the mark. The strict direction of the preceding paragraph governs how the pairs of each set combine. The reading converts no inconclusive outcome into support, and it leaves the refuter, the margin and the four labels as registered.

THE MARGINAL READING IS SCORED BESIDE THE ATTRIBUTION READING AND NEVER ALONE. The deciding unit makes the attribution of any measured excess to shared substrate a reading co-primary with the marginal one, because its own registered operating characteristics show the marginal reading firing under heterogeneity in fault difficulty shared by both panels with no substrate effect present. Both readings are therefore carried here. The sheet verdict follows the marginal reading exactly as registered above, and where the attribution reading returns not substrate-specific the refutation is reported as an excess in joint misses not attributed to shared substrate, with that reading and the unit's own registered characteristic printed beside the mark. A

refutation is not reported here without the attribution reading beside it, so that difficulty heterogeneity is never read as a substrate finding.

THE ANTI-CORRELATED CASE IS A FOURTH OUTCOME REPORTED IN ITS OWN RIGHT, and it is neither support nor refutation of this proposition as written, because the refuter above asks for an excess over the independence baseline and this is a shortfall against it. It is the reading where a pair lies wholly below the margin and no pair lies above it, and it says that correctors of the named population are systematically diverse in their blind spots on that fault set. Its consequence is registered now rather than described afterwards: under the equicorrelation model registered in the alternative baselines a negative common correlation returns a predicted correction elasticity above one half, so an anti-correlated panel points towards P6's refuter, which asks for a measured exponent interval lying entirely above one half. The outcome most favourable to stacking correctors is therefore the outcome that presses hardest on that bound, and the pressing is done by the correction-exponent estimator measuring the exponent, never by this unit inferring it from a dependence measurement. On the scoring sheet this outcome is therefore marked inconclusive, and the unit's own label is printed beside that mark rather than absorbed into it, so that the fourth outcome is neither lost from the sheet nor read as agreement.

TWO RULES OF THAT UNIT ARE CARRIED HERE BECAUSE THEY DECIDE WHAT COUNTS AS A PAIR AND WHAT COUNTS AS A KILL. The first is matched completeness: a pair's individual catch rates and its joint-miss rate are read on the same pairwise-complete set of items, so that a difference in which items two correctors managed to decide cannot present itself as a difference in how they fail together. The second is that the multiplicity correction is asymmetric by construction. The independence reading requires every primary pair to sit inside the margin and is valid at its nominal level without adjustment; the correlated and the anti-correlated readings each require only one pair to sit outside it, so both are read on intervals adjusted across the primary pairs by the Bonferroni rule at the familywise level that unit registers. Registering the asymmetry here stops the easier of the two readings being taken at the level fixed for the harder one, which is the manoeuvre that would otherwise make a kill cheaper than the support it is weighed against.

FALLS WITH IT: no numbered proposition. What is lost is the exactly-two-under-independence reading. Neither P6 nor P3 is lost with it. P3 survives because it reads the exponent measured on the system in front of it: correlation lowers that exponent and moves the boundary nearer one, which changes the value P3 is tested against rather than removing the test. P6 is the bound that this premise implies and not the premise itself, and the two move in opposite directions under the same evidence: a correlated result refutes this proposition while making that bound hold more comfortably, because correlation drives the correction elasticity below one half rather than above it. What is lost here is the equality at one half on which the value two depends, and the consequence is a boundary nearer one than two rather than no boundary at all. The conversion law and the co-scaling criterion are untouched.

HOW A FAILURE HERE FEEDS THE CEILING RATHER THAN MERELY DENTING IT, registered so the result is used and not only reported. The deciding unit measures pairwise joint misses against a conditional-independence baseline, from which the common pairwise correlation across the panel is estimable. That correlation enters the equicorrelation model registered in the alternative baselines

and returns a predicted correction elasticity, which the correction-exponent estimator then measures directly. Registering that chain in advance turns this proposition from a wager whose failure subtracts a premise into a measurement whose value is an input: a refutation supplies the number that lowers the predicted boundary, and the predicted and measured elasticities are then compared as a further check on the model that connects them. INSTRUMENT: DESIGN DRAFT, the correlated-failure unit. CONFIDENCE: moderate.

SUCCESSOR QUALIFICATION. THE DECIDING INSTRUMENT ON THE COUNTED ROUTE, NAMED BY SUCCESSOR VERSION 1.103. The instrument line above is retained as registered, and the correlated-failure unit it names, which the programme's register of units holds as Unit 08 (08-study-ah), remains this proposition's instrument as drafted; a verdict from it takes the counted route only where every term of its margin is a count, and is held by the gate otherwise. For a verdict on the counted route this successor extends to P17, the deciding instrument named here is the counted correlated-failure study, designed under the study identifier 08-AH-COUNTED as a counted variant of Unit 08. It is registered as its own unit, under the unit code its own registration carries, and names Unit 08's draft as its parent. Its status is that of a design draft until it is registered, and nothing of it runs before this successor and its own registration both read REGISTERED. In it this proposition is read on the quantity its refuter names, for each primary pair of correctors the log of the ratio of the observed joint-miss rate to the rate conditional independence predicts, on the faults both correctors decided, each decision recorded as a catch or a miss by frozen code against a label fixed in a sealed bank before any corrector reviewed any item, as the study's own registration specifies, with no judged term; after every mark the unit prints its domain in the words its own registration fixes, which name the counted measurement and state that no term is judged. Its correctors are reviewing checkers (line 474), each one engine under one registered review procedure, as the reading after the population paragraph above fixes, and the faults they review are of the checkable type (line 450), each fault's departure from a stated rule settled by frozen code under a typing codebook committed before any review, so that the fault set carries the type the refuter requires (line 1148). It carries no generator arm, both its panels being correctors reviewing an authored fault set, so it measures no overlap between a corrector's blind spots and a generator's (line 468), and its verdict carries the limit line 1142 registers for an authored fault population. Before a supported verdict is read, it shows from its own data that every primary pair could have returned a correlation of the registered size, as line 1150 requires. Its departures from the deciding unit described above are listed in its own registration, which governs, and include: two panels, each of one engine under several registered review procedures, in place of panels of distinct correctors sharing a substrate, with the pairs across the two engines carrying the substrate contrast; deciding fault sets of planted departures from the stated rules of a written charter, with a separate control set of technical faults that enters no mark (line 446); a calibration target of one half for the individual catch rate in place of the parent's band of 0.60 to 0.70, with the parent's detection floor replaced by a validity rule for each corrector and the demonstration just named; larger deciding sets, and a second fault set of the same construction on which a refutation must recur; replies cut short, blocked, empty or unreadable recorded as missing and never as misses; no generator arm; and the operator clause of the parent's rule on a separate system, from which the study departs because the author operates it. Condition A binds it as it binds every unit, through a per-stage attestation (line 420), and it records its regime, fixed-model inference with frozen code as the evaluator, before any

outcome is read (line 268). The external work this proposition stands against is engaged in the References field of this successor, as line 1509 requires before it is scored.

P18 · THE COMPOSITION OPERATOR PREDICTS THE FAMILY. Classified in advance by how a domain composes its inputs and its outputs, the scaling family that best fits that domain's data is the family the classification predicts, and not one chosen after the curve is seen.

WHY IT IS REGISTERED: the framework's structural theorem says the composition operator determines the functional form. That is the framework's claim to predict rather than to describe, and until now it appeared in this registration only as an assumption behind P1's model comparison.

WHAT MAKES THIS PROSPECTIVE RATHER THAN A CLASSIFICATION FITTED AFTERWARDS, which is the only way it carries any weight. A taxonomy flexible enough to be applied after a curve is seen will reproduce almost any set of families, so three things are frozen before any outcome is retrieved: the classification rule, written so that a third party applying it to a domain description returns the same family; the domain catalogue, named in advance and including domains whose curves the author has not seen; and the predicted family and exponent per domain, recorded with the rule that produced them. Domains whose data were already known to the programme are reported separately from domains classified before their data were obtained, and only the second group can support this proposition. The comparison is against two baselines, the empirical prevalence of each family across the catalogue and a flexible model free to choose a family per domain, because beating prevalence alone would show only that the commonest family is common. The dated artefacts that fix a classification rule and a set of per-domain predictions in advance are the commits made in the programme's public repository on 16 and 17 March 2026, whose attestation class is a public repository commit date corroborated by a server-side registry upload of the same folder rather than a registry server timestamp or an independent anchor; that catalogue was classified and sourced by this programme and has since been refitted by it with an added candidate family, so no claim of blindness is made for those domains here, the frozen-catalogue condition is met only by artefacts fixed after this registration, and freshness at domain level is established before any of those domains is entered as a group.

REFUTED BY: the predicted family failing to beat, by the margin registered with the deciding instrument and across the registered domain set, the comparator that belongs to the draw actually used: the balanced rate where the domains are drawn balanced across families, and the empirical prevalence where they are drawn unstratified.

WHICH COMPARATOR GOVERNS IS FIXED BY THE DRAW AND NOT BY PREFERENCE, and an earlier wording named prevalence alone. The deciding unit draws balanced and tests against one third, having identified and withdrawn its own earlier assertion about the base rate and recorded that an unstratified draw from its frame would inflate the apparent hit rate by about twenty percentage points before any predictive content was involved; scoring a balanced draw against prevalence would credit exactly that inflation.

AND THE SECOND REGISTERED BASELINE HAS NO COUNTERPART IN THE DECIDING UNIT: the flexible model free to choose a family per domain is not supplied there, which registers a

permutation null and a shuffled-outcome null instead. That is recorded here rather than left to be discovered, those two nulls stand in its place until a unit supplies it, and a verdict reached without it is reported as tested against the draw's own comparator alone.

FALLS WITH IT: the claim that the framework predicts form rather than fitting it afterwards. P1's form verdict is untouched, because a power law can hold in a domain whose family was mispredicted.

WHAT A SUPPORTED RESULT HERE WOULD LICENSE, and what it would not, stated in the restrictive form the licence rule above fixes: it licenses the statement that, classified in advance by how a domain composes its inputs and its outputs, the scaling family that best fits that domain's data is the family the classification predicts, on domains classified before their data were obtained, and it does not license any statement about the operator being causal, about domains outside the catalogue, or about the form verdict of the conversion proposition.

NOT EVALUABLE until the classification rule this proposition requires is shown to transmit to a third party. The condition above is that a third party applying the rule to a domain description returns the same family. A written procedure exists in the deciding unit as a prepared external-assessor packet and it has never been issued, so nothing measures whether a third party applying it returns the same family; until the registered agreement test reports at or above its threshold, the operator is reproducible in intention rather than in fact, and any classification made today would rest on the author's own reading case by case. That is the retrospective classification this proposition exists to exclude, so the honest status is that P18 cannot be scored, in either direction, before the agreement test reports and the rule is published with the domain catalogue. This is recorded as a limit on the proposition and not as a defence of it. INSTRUMENT: DESIGN DRAFT, BLOCKED ON A COMPOSITION-CLASSIFICATION RULE WHOSE THIRD-PARTY REPRODUCIBILITY IS UNMEASURED, the domain-catalogue units. The surrounding design is written and so is the rule: the deciding unit carries a written external-assessor packet, prepared and never issued, together with a registered agreement test comparing an external human classification against the frame's own locked class at a stated threshold. What is missing is the measurement and not the writing, and an earlier wording of this line said no such rule was written at all. This is a correction of the status and not a relaxation of it: until the agreement test reports at or above its registered threshold, the rule is not shown to transmit to a third party, and the five-value status vocabulary exists precisely to tell a runnable draft from a blocked one. CONFIDENCE: low.

P19 · EXTERNAL ALIGNMENT DOES NOT SCALE WITH CAPABILITY. Alignment achieved by mechanisms outside a system's own recursion does not improve as the capability of the system it governs rises: its exponent is indistinguishable from zero.

WHY IT IS REGISTERED: this is the headline prediction of the programme's own alignment paper, it is the premise of the argument that current practice cannot keep pace, and it was registered nowhere in this document.

REFUTED BY: on the zero-scaling axis, which is this proposition as worded, an alignment exponent whose interval lies entirely outside the registered equivalence region of plus or minus 0.10 about

zero, in either direction; and on the keep-pace axis, reported beside it, an interval lying entirely above one half.

THE ONE HALF ON THE SECOND AXIS IS A DECLARED CONVENTION AND IS REGISTERED AS ONE. An earlier wording called it this programme's own published falsification threshold and kept it for continuity with that. The provenance cannot be checked from this programme's registrations, where the nearest registered criterion for the same substantive question is a different number on a different scale, so the one half is registered here among the conventions the measurement field lists rather than offered as an authority a reader cannot reach.

PARITY ON THAT AXIS IS A SLOPE OF ONE, and one half is therefore the registered half-pace convention: a slope of six tenths clears the convention while alignment is still falling behind capability, and it is reported in those words and never as keeping pace. The keep-pace axis is not the sole refuter, and an interval at a third refutes the wording registered here while clearing nothing on the convention.

TWO VERDICTS ON TWO AXES, because the keep-pace threshold alone does not match the wording of the claim and the gap would otherwise be available afterwards. The claim says the exponent is indistinguishable from zero, and the keep-pace refuter fires only above one half, so an interval sitting precisely at a third would contradict the claim while refuting nothing. An earlier wording closed that gap with three outcomes and called them exhaustive. They were not: an interval can overlap the equivalence boundary at zero, overlap one half, lie materially below zero, or be wide enough to contain both zero and one half, and none of those four cases had a verdict. The repair is not a fourth outcome but two axes, which are jointly exhaustive because each is a partition.

THE ZERO-SCALING VERDICT, which is this proposition as worded: EQUIVALENT TO ZERO when the interval lies entirely within the registered equivalence margin of zero, fixed at plus or minus 0.10 on the exponent scale by the equivalence condition in the measurement field; MATERIALLY POSITIVE when it lies entirely above that margin; MATERIALLY NEGATIVE when it lies entirely below it; INCONCLUSIVE otherwise, including an interval that merely touches the margin.

THE MARGIN IS STATED ON AN EXPONENT SCALE AND THE DECIDING DESIGN MAY NOT BE, which is registered here because a margin carried across scales is the substitution this document forbids elsewhere. The plus or minus 0.10 above governs a unit that estimates an exponent and governs nothing else. The integrity-index scaling units drafted to decide this proposition estimate the slope of an operational integrity index against a capability score and register their own flat-equivalence bound on that regression scale, which is not the scale this document states its exponent margins on. The plus or minus 0.10 continues to bind any unit reporting on the exponent scale, which may narrow it and may not widen it, and a unit reporting on another scale does not widen it by reporting elsewhere. Where the deciding design registers a different estimand, its own registered margin governs its own verdict on its own scale, and the mapping from that estimand to the exponent this proposition is worded in is written down before any outcome is read. Where no such mapping is registered, that design's result is reported under its own name and this proposition stays NOT EVALUABLE rather than being scored on a substituted quantity.

THE KEEP-PACE VERDICT, read against the declared half-pace convention and reported beside the zero-scaling verdict: ABOVE THE THRESHOLD when the interval lies entirely above one half; BELOW THE THRESHOLD when it lies entirely below one half; INCONCLUSIVE otherwise.

NO DESIGN IN THIS PROGRAMME REGISTERS THAT RULE, so this axis is NOT EVALUABLE until one does, and the split verdict is not to be read as two live axes when only the first has an instrument. What would supply it is a unit estimating the same alignment quantity against capability and registering the half-pace threshold as its own decision rule before any outcome is read.

THE JOINT CELL IS THE RESULT, and it is reported in words rather than as a single label.

EQUIVALENT TO ZERO with BELOW THE THRESHOLD supports this proposition as worded.

MATERIALLY POSITIVE with ABOVE THE THRESHOLD refutes it on both axes and meets the keep-pace threshold.

MATERIALLY POSITIVE with BELOW THE THRESHOLD is the cell the author expects, and it is reported as external alignment scaling materially but not keeping pace: it refutes the indistinguishable-from-zero wording registered here, it does not meet the keep-pace threshold, it is never reported as support, and it leaves the programme's argument that external correction lags capability standing on the second axis while the first has failed. The confidence recorded below is stated for this proposition as worded, which is the zero-scaling reading, and is not a forecast of the joint cell: a result materially above zero but below the keep-pace threshold is anticipated here as a partial refutation of the wording registered, never as support for it.

MATERIALLY NEGATIVE is not predicted by anyone and is reported as a first-class outcome. Any cell containing INCONCLUSIVE is reported with that word on the axis that carries it.

AND WHERE NO DESIGN REGISTERS THE KEEP-PACE RULE, WHICH IS THE POSITION TODAY, THE JOINT CELL IS REPORTED WITH NOT EVALUABLE ON THAT AXIS: the zero-scaling verdict alone carries this proposition's mark on the scoring sheet, and the missing rule is printed beside the mark. The four cells above are the rule that applies once a keep-pace design exists, and not one of them can be entered until one does, since each of the four requires a keep-pace verdict that nothing can currently return.

FALLS WITH IT, SPLIT BY AXIS BECAUSE THE TWO AXES COST DIFFERENT THINGS: failure on the zero-scaling axis removes the claim that external alignment is practically flat, and nothing more. What removes the programme's argument that external approaches cannot keep pace, and with it the motivation for the protocol, is a result clearing the keep-pace threshold on the second axis. The cell the author expects, materially positive but below that threshold, therefore costs the flatness claim and leaves the lagging claim standing; an earlier wording charged both to either axis and so overstated what a partial refutation takes. The three laws are untouched, which is why it is registered separately from them. INSTRUMENT: DESIGN DRAFT, BLOCKED ON A REGISTERED MAPPING FROM THE INTEGRITY-INDEX REGRESSION SCALE TO THE EXPONENT SCALE, the integrity-index scaling units, named here for what they measure rather than for the quantity they would need a mapping to reach. They estimate the slope of an operational integrity index against a capability score and register their own flat bound on that regression scale, and their own text says

the estimand is not an exponent margin; an earlier wording called them the alignment-exponent units, which named them for the one scale this instrument line exists to say they do not report on. The units are drafted; the mapping this proposition's own rule requires before their result can be scored against it is not, and the status names that rather than implying a runnable verdict. Those units bear on the zero-scaling axis alone. The keep-pace axis has no deciding design registered anywhere in this programme and is NOT EVALUABLE, stated here beside the other axis so that a split verdict is not read as two live axes when one of them has no instrument. CONFIDENCE: the author states it as a sentence rather than as a rung of the scale above, and his sentence is carried here in his own words: "I believe it is high probability that external alignment does not scale with capability".

SUCCESSOR QUALIFICATION. P19 remains a survey in no register. Its refutation touches none of Laws I to V; the historical sentence about three laws is therefore retained with that wider, explicit no-carrier reading.

P20 · THE CEILING RELATION TAKES THE CORRECTED FORM. Where a stability boundary can be located and a correction exponent measured on the same system, the boundary follows one over one minus the correction exponent, rather than one over the correction exponent or a fitted constant.

WHY IT IS REGISTERED: this programme corrected the reciprocal form on 16 August 2026 and replaced it with the relation registered here. The two agree at exactly one half, which is the programme's headline value, so the replacement has never been tested anywhere the forms separate. P3 assumes the corrected form; nothing until now tested it. Registering the discrimination is the difference between a correction that was reasoned and a correction that was measured.

THE RIVAL SET IS WIDER THAN THREE, and naming it now is the difference between a test and a confirmation. The candidates are the registered relation; the retracted reciprocal, kept so that the correction of 16 August 2026 is itself exposed to defeat; a fitted constant, which is what a frontier that does not depend on the corrector would look like; the burden-intensity generalisation, in which the correctable burden generated per unit of capability changes with capability and the frontier becomes one over the quantity one plus that elasticity minus the correction-leverage exponent; a smooth model with no sharp boundary at all, which is what a reader who doubts that stability has a threshold would propose; and a two-rate feedback model in which the system moves between a fast improving state and a slower correcting one under a fixed switching rule, which is named here because it is the most credible way to produce what looks like a boundary without there being one. It is separated from the registered relation by an observation rather than by an argument: a two-rate model drives the correction-to-burden ratio towards a bounded fixed point, approached geometrically and from one side, while a threshold model carries the ratio through zero and onwards with the sign changing at a location the relation predicts in advance. The deciding unit registers the geometric-approach test and the crossing test together, and a ratio that flattens towards a positive floor is scored for the two-rate model however far it has fallen. The registered relation is preferred only where it beats every one of them on held-out systems by the registered margin.

A FOURTH OUTCOME, NOT DISCRIMINATING, because the forms agree where the programme's headline sits. The registered relation and the retracted reciprocal return the same value at exactly

one half, which is the value this programme is known for, so cells clustered there cannot separate them however many there are. Where the realised correction-leverage exponents do not clear the deciding unit's registered admissibility margin away from one half, the verdict is NOT DISCRIMINATING and no preference is reported in either direction. What the comparison needs is distance from one half rather than cells on both sides of it, because the two forms separate wherever the exponent is far enough from the value at which they agree, and a both-sides requirement would refuse a verdict on evidence that in fact discriminates. That outcome is neither support nor refutation and is expected to be common, because engineering a corrector to a target exponent is the hard part of this experiment rather than an incidental detail of it. On the sheet it is marked inconclusive, with the reason for it printed beside the mark.

A FIFTH OUTCOME IS IMPORTED FROM THE DECIDING UNIT RATHER THAN OVERRIDDEN, because that unit returns an outcome this entry did not carry. It registers a minimum count of admissible cells on each side of one half and returns INDETERMINATE by construction where either side falls short, giving its own reason: the two forms make the same prediction reflected about one half, so a sample lying wholly on one side cannot separate them. That outcome is registered here under the deciding unit's own name, is marked inconclusive on the sheet with the unit's label printed beside the mark, and is never read as a preference in either direction.

WHICH DOCUMENT GOVERNS WHERE THE TWO RULES DIFFER, said because the conflict is live rather than hypothetical. This registration governs the verdict this proposition takes; the deciding unit governs whether its own run is admissible and returns its own label. A run that unit declares indeterminate therefore yields no verdict here, whatever this entry's distance rule would otherwise return. The reason this matters is registered too: this programme's own conjecture that the correction exponent for same-class correctors sits at or below one half predicts that the far side may be hard to populate, in which case the form instrument is silent by construction, and the silence is reported rather than worked around.

REFUTED BY: the discrimination registered with the deciding instrument selecting a rival over the registered relation across a majority of admissible cells, on systems held out of the fitting.

DEPENDS ON: the crossed variation that identifies the correction elasticity, and not on P1. An earlier version of this entry said a boundary written in the growth exponent has no subject where no growth exponent exists in the tested class, and cited the dependency map as stating it. The map's own sentence says the opposite, and states the same dependency in its own words rather than in these; this entry's own statement of that dependency is that this proposition depends on the crossed variation that identifies the correction elasticity and not on P1. The reason the map is right and the earlier entry was wrong: this proposition reads the correction elasticity, which is not recoverable from a single observed path, so it stays NOT EVALUABLE until the crossed variation the direct boundary-mapping unit is designed to supply exists, and that holds whether P1 holds or fails.

FALLS WITH IT: P3's derivation, and the ceiling's standing as a derived quantity rather than a fitted one. P6 survives. The registered relation does not keep its standing as the best available reading in that event: where the refuter selects a rival, that selection is the displacement, and the rival is reported as the better-supported form.

THE FORM COMPARISON CARRIES ITS OWN JOINT UNCERTAINTY, on the rule P3 states for its paired difference and for the same reason. The boundary exponent and the correction-leverage exponent are both estimated, each candidate form is a non-linear function of the second, and the error in a form's predicted boundary is therefore neither symmetric nor independent of the error in the boundary measured beside it. The comparison between forms is run through the joint bootstrap or measurement model the deciding unit freezes before any cell is fitted, propagating the covariance of the two estimates and the curvature of each form. Computing an interval for the measured boundary, computing a second interval for each form's prediction and preferring whichever pair sits closest is not admissible here, and neither is treating the correction-leverage exponent as a known input while the forms are compared. The burden-intensity generalisation is held to the margin P3 already states rather than to a looser one: the registered relation is preferred over it only where the burden-intensity elasticity is practically equivalent to zero under the registered margin and the simpler form predicts held-out frontiers at least as well, and a generalisation that is merely not preferred is not a generalisation shown to be unnecessary.

TWO FURTHER INSTRUMENTS READ THIS FORM AND NEITHER DECIDES IT. The paired frontier-and-exponent unit registers the same contrast on its own cells, setting the registered relation against the retracted reciprocal and against a threshold in the correction exponent alone that does not move with the growth exponent, and repeating that contrast on the cells that survive an exclusion around the boundary where the forms agree. The direct boundary-mapping unit, where the decisions blocking it are resolved and it runs, reads the same form as a location: on held-out confirmatory cells the zero crossing of the correction-pressure exponent, measured on its own stream rather than read off the fit it is scored against, is predicted to fall where the registered relation puts it, and it is compared there against that unit's own frozen rival set. Both corroborate and neither substitutes for the form-discrimination unit's verdict. Agreement between the three is agreement inside one programme, which shares its design assumptions and its scale conventions, and it is reported in those words rather than as replication.

THE RIVALS ARE ALSO TESTED JOINTLY AND NOT ONLY ONE AT A TIME, because two omitted terms of opposite sign can cancel and leave the simple relation looking correct for the wrong reason. The general form recorded in the alternative baselines carries a burden-intensity elasticity and a direct depth dependence of correction service at once, and the registered relation is the case where both vanish. The deciding unit therefore fits the combined model as well as each one-parameter departure, and the simple relation is preferred only where the combined fit does not beat it on held-out pressure by the registered criterion. A one-at-a-time comparison that each departure fails is not evidence that the joint departure fails.

A FURTHER RIVAL, FROM THE PROGRAMME'S OWN GENERAL FORM, and the scale on which the comparison is valid. The general boundary recorded in the alternative baselines carries a direct depth-dependence of correction service, and where that dependence is not zero the boundary is one plus it, divided by the correction shortfall. The relation registered here is the case where it is zero. That one-parameter generalisation joins the rival set as a named alternative, and where the deciding unit can estimate the dependence it reports both. The comparison between forms is valid only on the capability scale condition B fixes, for the reason given under the same-class bound: a rescaling

changes the numbers a form comparison runs on without changing the system. INSTRUMENT: DESIGN DRAFT, the form-discrimination unit. CONFIDENCE: high.

P21 · IN-LOOP CORRECTION PULLS AWAY WITH DEPTH. Correction that participates in each subsequent revision round outperforms correction that sees only the finished output, and the advantage widens as recursive depth rises rather than staying constant.

WHAT THIS PROPOSITION MAY AND MAY NOT BE READ AS MEASURING. The arms share their initial conditions and are then allowed to diverge, because a design that holds the trajectories together measures something other than what is registered here. An effect observed along a realised path does not by itself identify the direct dependence of correction service on capability: a path carries the allocation and the trajectory together, so a difference between the arms is evidence about where correction is placed and not an estimate of that dependence, which the chain rule in the variable relationships keeps separate. A supported verdict here is a statement about placement, never a measurement of an exponent.

THE TITLE NAMES WHAT THE APPARATUS VARIES, which an earlier title did not. The deciding design varies whether correction enters the next round or only inspects the output; it does not manipulate substrate, composition class or failure-mode independence, so a supported result establishes an access interaction and not that physical or architectural embedding caused it. The interaction is the claim: a constant advantage at every depth would be a different and weaker result and is not what the protocol asserts.

WHY IT IS REGISTERED, AND WHY ITS ABSENCE WAS THE LARGEST GAP IN THIS SET: this is the claim the proposed protocol exists to justify, and the instrument that would decide it is the embedded-versus-external correction designs, named in this entry's own instrument line below. An earlier wording reached that instrument through P15's line instead, which no longer names it. The set carried the protocol's premise, on the class of current practice, and three of its four carriers, on build order, on corrector composition and on the decay of externally installed gains, and none of them is the depth interaction. A registration that names an instrument while omitting the proposition that instrument was built to test leaves its central engineering claim unregistered beside its own evidence.

SUCCESSOR QUALIFICATION. The protocol this paragraph calls proposed is the Eden Protocol, as this successor names it at line 796; this entry's registered words stand unchanged.

REFUTED BY: no interaction between correction placement and depth beyond the tolerance registered with that instrument, in a design whose own sensitivity shows an interaction of that size would have been detected. A design too small to find the interaction is a failed instrument and not a refutation, as condition D already provides.

NOT SATISFIED BY A CONSTANT ADVANTAGE: an embedded arm that beats the external arm by the same margin at every depth is reported as a level difference and is scored against this proposition, not for it, because a level difference is consistent with embedded correction simply being better implemented and says nothing about recursion.

SUCCESSOR QUALIFICATION. THE DIRECTIONAL READING, ADDED BY SUCCESSOR VERSION 1.103 BEFORE ANY DATA THAT COULD DECIDE THIS PROPOSITION. The interaction claimed here has a direction, the advantage of correction inside the loop widening as depth rises, while the refuter above names only the absence of an interaction beyond the tolerance registered with the deciding instrument. This document's general rule scores as refuted an interval lying entirely in the region a prediction excludes (line 512), and for this proposition that region is every interaction short of a widening of the registered size. With the tolerance taken on either side of zero, two results are therefore scored against this proposition, as the preceding paragraph scores a constant advantage: an interval for the interaction lying wholly below the positive tolerance, in a design whose own sensitivity shows that an interaction of the tolerance's size would have been detected, which shows no widening of the registered size; and an interval lying wholly below the negative tolerance, which shows a narrowing. Read alone, the refuter above reaches such an interval only where it lies wholly within the tolerance; this reading states the other cases. It converts no inconclusive outcome into support, and it leaves the refuter, the tolerance and the margin as registered. An interval that straddles the positive tolerance stays inconclusive, and so does one lying wholly below it in a design without that sensitivity, unless it lies wholly below the negative tolerance.

WHAT THE DECIDING UNIT VARIES, in its own terms, because it is narrower than this proposition sounds. That unit crosses three correction regimes with recursive depth on identical artefact trajectories: an external monitor that sees the finished artefact and may reject it, an embedded arm whose correction participates in the next round, and a no-correction control, matched on model, calls and token budget, with blinded evaluators and a capability-cost check. Its own registration is explicit that what it manipulates is information access, whether the corrector sees the process or only its output, and not composition class or failure-mode independence. This proposition is therefore decided on the access axis and its verdict is reported with that scope attached.

IT ALSO VARIES THE ACTION SPACE, WHICH IS ONE MORE THING THAN INFORMATION ACCESS, and the limitation is registered beside the scope. The external arm may only accept or reject, and on a rejection the previous round's artefact is carried forward; the embedded arm generates under constraint and can therefore reach artefacts the external arm's action space does not contain. At any material rejection rate a widening advantage accumulates with rounds from that difference alone, whether or not the monitor's view degrades, so a supported verdict can be produced by the action space rather than by the information asymmetry this proposition names. The deciding unit therefore reports the per-round rejection rate beside the interaction, and a widening advantage arriving without the registered fall in the monitor's detection rate as artefact complexity rises is reported as unexplained by this framework and never as support for the information-asymmetry mechanism, which is the conjunction this document imposes on the corrector-class proposition for a claim of the same shape.

SUCCESSOR QUALIFICATION. TWO VERDICTS, ADDED BY SUCCESSOR VERSION 1.103 IN THE FORM REGISTERED FOR THE CORRECTOR-CLASS PROPOSITION (line 1080). This proposition's sheet mark follows the interaction, which is the proposition as worded: supported, refuted, inconclusive or untested by the rules of this entry. Its mechanism axis is reported beside the mark and never in its place. That axis asks whether the monitor's detection rate falls as the artefact it

inspects grows, which is the registered fall named above, and whether the part of the widening that fall can carry clears the tolerance once the part owed to the outside arm's action space is separated out, by a method the deciding unit registers before any outcome is read. It is reported supported, unsupported, inconclusive or NOT EVALUABLE, and its terms are taken as the mark's are: where every term of the axis is a count settled by frozen code against ground truth fixed before the run, it is reported beside the mark with the domain, and where any judged term enters it, the gate binds it in full and it is reported NOT EVALUABLE, with the gate named beside it, until the gate passes. A supported mark beside an unsupported mechanism axis is reported in those words, and wherever no fall is registered the preceding paragraph's words apply: the widening is reported as unexplained by this framework and never as support for the information-asymmetry mechanism.

THE TRADE-OFF THIS PROPOSITION DOES NOT RESOLVE, stated so that a supported result is not read as resolving it: embedding correction in the substrate that computes maximises the corrector's scaling with capability and minimises its independence from the generator, because a corrector built into the system shares the system's blind spots by construction. Those are the two axes named in the admissibility conditions above, and embedding moves them in opposite directions. Whether the gain on scaling outweighs the loss on independence is not registered here and is not decided by this proposition. A depth interaction in the embedded arm's favour is evidence on the scaling axis alone, and the arrangement that would settle the trade-off, an arm combining embedded correction with an independent external layer, is not registered by this proposition and is not claimed by it.

FALLS WITH IT: the protocol's engineering case, jointly with P12, P13 and P15. The three laws are untouched, which is why it is registered separately from them.

SUCCESSOR QUALIFICATION. The three untouched laws here are Laws I, II and III. Under successor version 1.103 this proposition is the placement limb of Law V, so a refuted P21 refutes Law V. It refutes neither Law IV nor P15, and no other numbered proposition is newly made to fall.

WHAT THIS PROPOSITION IS IN THE FRAMEWORK'S OWN QUANTITIES. The general boundary is set by four quantities and this proposition is a claim about one of them: that correction embedded in the recursion carries a larger direct depth-dependence than correction applied from outside it. Where that holds, the boundary rises with it, which is the engineering proposal's mechanism written as a single registered parameter rather than as a preference between architectures. The path elasticity of correction service is measurable per arm, so the quantity the claim is about can be reported directly beside the outcome the unit scores. This widens nothing: the registered claim, its verdicts and its margin are unchanged, and the limitation that the deciding unit varies information access rather than composition class stands.

THIS PROGRAMME HAS ALREADY ATTEMPTED A WEIGHT-LEVEL VERSION OF THIS CONTRAST TWICE, neither attempt discriminated, and both are disclosed here rather than left for a reader to find. A published experiment of this programme, Paper VIII, The Load-Bearing Test, of 18 March 2026, revised 2 September 2026, OSF DOI 10.17605/OSF.IO/7YJ4E, fine-tuned a three-billion-parameter instruction-tuned model under three loss functions, one of them an entangled arm standing for embedded correction. It was run twice: first at nine training examples over eight layers for one hundred iterations, then at two hundred and ninety-five examples over sixteen layers for five

hundred iterations. Both runs produced catastrophic forgetting. Every fine-tuned condition scored worse on capability than the unmodified base model, which scored 7.68 against 4.00 for the best of them in the larger run. That paper's own reading is that the base model's existing preference training is too strong for low-rank adaptation on a few hundred examples to improve rather than degrade it, and it names what a discriminating run would need: several thousand training examples, a larger model, a base model without preference training, or full fine-tuning rather than low-rank adaptation. The same paper's behavioural experiment returned a null, its three conditions statistically indistinguishable, and its one positive result is a simulation rather than a system, which makes it prediction-side and never an empirical outcome.

THE STANDING OF THAT RESULT AGAINST THIS PROPOSITION, STATED EXACTLY. It is inconclusive and it is not adverse, and the difference is one this document insists on elsewhere: all three arms degraded, so the contrast registered here was never put, and a design that fails to measure anything is not a design that found no effect. Nor does that paper test what this proposition claims. It asks whether embedded safety costs capability; this proposition asks whether the advantage of embedded correction widens with recursive depth, which no arm of it varied. It is recorded because the scale-up was already tried and failed the same way, which is worth more to a reader than the first attempt alone, and because the author's nearest attempt at his own engineering claim returning nothing is exactly the fact a registration disclosing only its untried hypotheses would omit. That paper's own conclusion is that whether embedded safety produces measurable benefit remains open. INSTRUMENT: DESIGN DRAFT, the embedded-versus-external correction designs. That unit's registered sensitivity was measured on a smaller design than the one registered, and its negative-slope hypothesis has no operating characteristic; both are recorded against that unit and are closed before this proposition is scored. CONFIDENCE: high.

SUCCESSOR QUALIFICATION. THE DECIDING INSTRUMENT ON THE COUNTED ROUTE, NAMED BY SUCCESSOR VERSION 1.103. The instrument line above is retained as registered, and the designs it names remain this proposition's instrument for any verdict whose margin has a judged term, under the gate. For a verdict on the counted route this successor extends to P21, the deciding instrument is the counted embedded-versus-external correction study, designed under the study identifier 41-D-COUNTED as a counted variant of Unit 41 (41-study-d). It is registered as its own unit, under the unit code its own registration carries, and names Unit 41's draft as its parent. Its status is that of a design draft until it is registered, and nothing of it runs before this successor and its own registration both read REGISTERED. In it this proposition is read on one endpoint: the per-round value of each arriving unit of work, plus one for a sound unit committed, zero for none and minus one for a faulty unit committed, each settled by frozen code against ground truth fixed in a sealed bank before the run, as the study's own registration specifies, with no judged term. The interaction is the slope across depth of the value of the arm whose correction enters each round minus the value of the outside arm, read against the tolerance the unit registers, and after every mark the unit prints its domain in the words its own registration fixes, which name the counted measurement and state that no term is judged. The two arms its mark compares are both reviewing checkers, and the faults they act on are checkable (lines 450 and 474); neither is the embedded correction mechanism of the engineering proposal, and no verdict in it is a verdict on that mechanism. Its departures from the deciding unit described above are listed in its own registration,

which governs, and include: counted scoring by frozen code against a sealed bank in place of blinded evaluators; arrivals administered from a sealed bank with planted faults rather than generated; an arm whose correction enters each round by repairing the arriving draft rather than by generating under constraint; depth as the count of administered arrivals, none revised twice, so that depth and the size of the artefact in the outside arm's view rise together, and the unit does not separate a monitor's loss of detection on a longer artefact, which a ladder of artefact sizes with no loop would also show, from any effect of recursion as such, while what the loop alone adds is the outside arm's stock of earlier misses, which the unit's split assigns to the rival of line 1258 and not to the mechanism; a matched control call that reviews the outside arm's cleaned view; a token budget matched as an allowance, with the tokens used reported; and its own registered tolerance beside the parent's drafted margin. Condition A binds it as it binds every unit, through a per-stage attestation (line 420); the bounded-endpoint clause applies to its endpoint (line 434); and it records its regime, fixed-model inference in an administered revision loop, before any outcome is read (line 268). The two items the instrument line above records against the parent are met in this unit before this proposition is scored on it: its sensitivity is demonstrated on the design it registers, and its outside arm's own slope carries a registered operating characteristic. The external work this proposition stands against is engaged in the References field of this successor, as line 1509 requires before it is scored.

P22 · CHECKABLE CORRECTION CARRIES THE HIGHER EXPONENT, SO THE CEILING IS LOWEST WHERE THE SAFETY CLAIM LIVES. The checkable correction-leverage exponent exceeds the judged correction-leverage exponent on the same systems under the same instrument, and since the third law returns a ceiling rising with the exponent, the ceiling is strictly lower for judged correction than for checkable correction.

THE INEQUALITY IS STRICT, and the decision rule is on the difference rather than on the two exponents separately. An earlier wording said at least, which would have let equality satisfy the proposition while equal exponents give equal ceilings and the stated consequence would not follow.

FOUR OUTCOMES, because a strict positive ordering fails in two scientifically different ways and collapsing them would hide which one occurred.

THE MARGIN THESE FOUR OUTCOMES ARE READ AGAINST IS THE ONE THIS DOCUMENT ALREADY REGISTERS FOR EVERY EXPONENT COMPARISON, and that is said before them rather than after. All four turn on an equivalence margin fixed by the equivalence condition in the measurement field. Both quantities compared here are correction-leverage exponents, so that margin is 0.10 on the exponent scale, held in common with the other exponent comparisons so that this proposition is not judged on an easier scale than they are. An earlier wording left the number to the author and reported this proposition as waiting on it, when the number was already registered for every comparison of this kind; applying it here is a correction and not a new margin. A deciding instrument may register a smaller margin and may not register a larger one. This proposition is NOT EVALUABLE for want of an instrument, which is the status its instrument line carries, and no longer for want of a number.

A MINIMUM DETECTABLE DIFFERENCE IS REGISTERED WITH IT, for the reason the same-class bound already gives when it returns inconclusive on a survey that lacked the precision to have found a counterexample of the registered size: the quantity scored here is the difference of two exponents estimated by the same feeder that bound reads, and on such a difference the practical-equality outcome may be unreachable at any true value, which would make inconclusive the modal outcome of a run built on today's precision. The deciding instrument therefore registers, before any outcome is read, the smallest difference it could detect at the registered margin, and reports it beside the verdict. A proposition whose rule is stricter than its neighbour's must not be the one that omits that arithmetic.

CHECKABLE HIGHER, scored SUPPORTED, when the interval on the checkable exponent minus the judged exponent lies entirely ABOVE THE EQUIVALENCE MARGIN, not merely above zero, and the ordering replicates on fresh data at the same settings.

JUDGED HIGHER, scored REFUTED, when the interval lies entirely below the negative of that margin and the reversal replicates.

PRACTICALLY EQUAL, also scored REFUTED, when the interval lies entirely inside the registered equivalence margin around zero on the exponent scale, fixed by the equivalence condition in the measurement field.

THE MARGIN AND NOT ZERO IS THE BOUNDARY ON ALL THREE, corrected here for the reason given under P8: a difference interval sitting just above zero but inside the margin would otherwise have satisfied the support rule and the practical-equality refutation at the same time. No later short refuter phrased as mere negativity overrides this rule either: an interval lying below zero but inside the margin is practical equality and is scored under that outcome rather than under the reversal. INCONCLUSIVE otherwise, reported in that word and never as support for a mechanism that would then not have shown itself. Both the reversal and the practical equality refute the strict ordering this proposition asserts, and they are reported separately: one says the ceiling orders the other way, the other says it does not order at all.

WHY THE FRAMEWORK ALREADY PREDICTS THIS, THOUGH IT HAS NEVER SAID SO: the bound rests on corrections accumulating as though independent, and independence is not a property of correction in general. It is a property of how far the corrector's failure modes overlap the generator's, and that varies with what is being corrected. A checker that can appeal to something outside itself, a recomputation or a stated constraint, fails in ways largely unlike the generator's, so it can sit outside the generator's class. A model judging another model's conduct shares a training distribution with it, so its blind spots are correlated by construction and it cannot. The ordering follows from the framework's own premise rather than from any new assumption, and it is sharper than a single value of two because it says where the value sits and where it does not.

THE OVERLAP IS THE MECHANISM AND NOT A PROOF, and the difference matters to what a null would mean. Lower failure-mode overlap does not entail a larger elasticity of correction against capability: the premise makes the ordering expected, and the measurement decides it. A PRACTICALLY EQUAL result would therefore not show that the reasoning above is incoherent; it

would show that the overlap difference does not reach the exponent, which is a finding about the mechanism and is reported as one.

AND IT IS LOWEST FOR JUDGED CORRECTION, which is where this programme's safety claim lives. That is registered plainly because it is the least convenient consequence of the framework for the framework, and because a reader who works it out unaided will assume it was avoided.

REFUTED BY: a measured judged correction-leverage exponent exceeding the checkable one on the same systems under the same instrument, with the reversal replicating on fresh data at the same settings.

WHAT THIS PROPOSITION IS IN THE FRAMEWORK'S OWN QUANTITIES, and why it decides more than its place in this list suggests. It is the claim that the correction elasticity is larger for checkable correction than for judged correction, which is the second of the two correction quantities among the four an intervention can move. Its consequence follows from the complexity argument recorded above: where a class of fault admits a decision procedure it is excluded by construction rather than sampled for, so for that class the elasticity is not limited by the square root of effort and the boundary written on it need not be finite. Judged correction has no such route available to it. This proposition decides the ordering of the two typed exponents.

IT DOES NOT BY ITSELF DECIDE WHETHER A FINITE BOUNDARY APPLIES, and an earlier wording said that it did. The ceiling relation returns a finite value only where the correction exponent lies below one, and an ordering between two exponents fixes the absolute position of neither: both may sit below one, both above it, or one on each side. The regime question is therefore conditional on the relation's domain, on the other model conditions registered with it, and on the absolute estimates, and it is reported separately from the ordering this proposition actually tests. This entry is a correction of an earlier one that named the correlated-failure unit, and the correction matters more than the change of name: that unit estimates the excess of joint misses over a conditional-independence baseline across a panel of correctors, which is a dependence between failures and is not an elasticity of correction against capability. An instrument measuring correlation cannot decide an inequality between two exponents. What would decide this proposition is the correction-exponent estimator, which fits corrective strength against overseen capability per corrector family, run over two matched fault populations of stated type on the same capability ladder, with the two exponents estimated under one instrument so their difference carries a joint interval. That extension is not built, does not exist anywhere in this programme's own work, and will be registered as its own unit or as a recorded amendment to the estimator's own registration before any outcome is examined. The correlated-failure unit remains named here as mechanism evidence: it can show whether the failure-mode overlap this proposition's reasoning invokes differs by type, and it cannot supply the exponents.

FALLS WITH IT: the reading of the value two as a single number that holds across correction types. P6 survives as a statement about whichever type its own instrument measures, and the ceiling relation is untouched. INSTRUMENT: NONE, and a typed extension of the correction-exponent estimator is required. It does not exist. CONFIDENCE: moderate.

THE SCORING SHEET, so that scoring this registration needs nothing but this registration. Mark each line supported, refuted, inconclusive or untested. Those four are exhaustive, and untested is not a polite form of supported.

NOT EVALUABLE IS THE REPORTED FORM OF UNTESTED WHEREVER A PROPOSITION WAITS ON SOMETHING NOBODY HAS SUPPLIED, and the classes it covers are named here rather than shown by one example. They are: a missing upstream measurement; a rule the proposition's own verdict requires and no design registers; a number the author holds and has not yet recorded in the deciding instrument; an instrument that does not measure the endpoint the proposition is worded in; two arms that cannot be placed on one coordinate; a completed run whose registered regime condition was never reached; and per-item outputs another team has not released. The label appears in the bodies of eighteen of the twenty-two, and for several of those it is the present status of the whole proposition rather than one outcome among four. It is scored as untested, so the four marks above stay exhaustive, and what is missing is printed beside the mark: the quantity where the blocker is a quantity, and the rule, the number, the endpoint or the output where it is not. Each proposition's own body names which of them is missing. Then apply the takes-with-it column and read off what remains. A reader who finds the columns below disagreeing with the prediction bodies above should trust the bodies and treat the disagreement as a defect to be reported.

ID THE CLAIM, COMPRESSED SCOPE CONFIDENCE INSTRUMENT TAKES OUTCOME P1 power law and depth advantage S1 high DESIGN DRAFT P2 [ ] P2 sustainable profile turns over S1 moderate DESIGN DRAFT nothing [ ] P3 relation upper-bounds frontier S2 low DESIGN DRAFT nothing [ ] P4 correction out-scales drift S2 moderate DESIGN DRAFT unscopes P2 P3 P16 [ ] P5 exponent derived, not fitted S1 not stated DESIGN DRAFT nothing [ ] P6 correction exponent bounded S3 low DRAFT, BLOCKED nothing [ ] P7 deployed correction is weak S6 moderate DRAFT, BLOCKED nothing [ ] P8 exponents exceed the null S1 moderate DESIGN DRAFT nothing [ ] P9 the cross-domain form S4 low DRAFT, BLOCKED nothing [ ] P10 scorer-family dependence S6 moderate DESIGN DRAFT nothing [ ] P11 the two not interchangeable S1 not stated DESIGN DRAFT nothing [ ] P12 build order moves the exponent S7 low NONE nothing [ ] P13 same-family corrects worse S5 moderate NONE nothing [ ] P14 gains shrink when blinded S5 moderate NONE nothing [ ] P15 external gains decay S5 moderate NONE nothing [ ] P16 the prohibition itself S2 mod to high DRAFT, BLOCKED P3 [ ] P17 panel failures are independent S3 moderate DESIGN DRAFT nothing [ ] P18 operator predicts the family S4 low DRAFT, BLOCKED nothing [ ] P19 external alignment is flat S6 his words DRAFT, BLOCKED nothing [ ] P20 ceiling takes corrected form S2 high DESIGN DRAFT nothing [ ] P21 in-loop pulls away with depth S7 high DESIGN DRAFT nothing [ ] P22 ceiling orders by fault type S3 moderate NONE nothing [ ]

TAKES lists numbered propositions only. What each proposition costs the programme when it falls, which is a different column, is written out in the dependency map above and is not compressible to a phrase. The five propositions with no instrument are P12, P13, P14, P15 and P22. The INSTRUMENT column is a deliberate abbreviation of the body lines and adds no status. Its entries stand for three of the five registered values: NONE; DESIGN DRAFT; and DESIGN DRAFT, BLOCKED, with the blocker named, which the column cannot hold at this width. Each proposition's own INSTRUMENT line above prints the value in full and names the blocker where the value is

blocked, and names the deciding unit by function wherever one is named. P4 also unscopes P22's ceiling consequence, which is part of a proposition rather than a whole one and so does not fit this column; a scorer meeting a failed P4 marks that consequence untested along with P2, P3 and P16.

P20 takes P3's derivation, which is likewise part of a proposition rather than a whole one and so does not fit this column either; the dependency map above and P20's own line carry what a refuted P20 costs P3.

P5 returns a pair of words rather than one. Its sheet mark follows the prediction result and the identification word is printed beside the mark, so a supported prediction returned with NOT IDENTIFIED is marked supported with that word beside it and is never read as establishing the mechanism.

P6 admits no supported verdict. Its registered outcomes are refuted, no counterexample found among N, and inconclusive; the absence of a counterexample is never marked as support, so supported never applies to that line. A survey returning no counterexample is marked inconclusive with the number examined printed beside the mark, and refuted and inconclusive carry their ordinary meanings.

P19's confidence column reads his words because the author stated that confidence as a sentence rather than as a rung of the scale, and the sentence stands in P19's own entry.

P16's confidence column reads mod to high because the column is set to a fixed width that his words, moderate to high, do not fit; his words stand whole in P16's own entry and nothing in this sheet narrows them.

P8 returns a second verdict beside the first. Its sheet mark follows the first verdict and the CLEARS UNITY result is printed beside the mark, so a supported first verdict beside a failed CLEARS UNITY is marked supported with those words beside it and is never read as support for the conversion claim.

Eleven lines are held by the scoring gate registered above and are marked here because this sheet has no column for it: P3, P4, P6, P13, P14, P16, P17, P19, P20, P21 and P22. Until the alignment scorer passes both validity tests, a result offered against any of those eleven is marked untested on this sheet whichever way it points, except where the scoping under condition E1 records that no term in the verdict's margin is judged, in which case the domain is printed beside the mark; and the author's own delayed re-scoring does not discharge that gate for any of them.

SUCCESSOR QUALIFICATION. Under successor version 1.103 the exception in the preceding sentence also reaches P21: where the scoping this successor adds beside condition E1's, in the Measurement and Data field, records that no term in P21's margin is judged, P21's mark is entered with its domain printed beside it. The exception reaches P17 in the same way, through the scoping this successor extends to P17's margin under condition E1: where that scoping records that no term in P17's margin is judged, P17's mark is entered with its domain printed beside it. P21 also returns a second verdict beside its mark, as P5 and P8 do: the sheet mark follows the interaction, and the mechanism axis that P21's entry registers is printed beside the mark and never in its place.

WHAT THE SHEET SHOWS AT A GLANCE, STATED SO THAT IT CANNOT BE READ SELECTIVELY. Three propositions carry high confidence, one of which is the frame rather than a finding, and a fourth is held across the top two rungs. Five have no instrument in existence: P12, P13, P14, P15 and P22. P22 lost its instrument in an earlier version when the unit named against it was found to measure a different quantity, P5 gained one on 5 September 2026 when the design its entry specified was drafted as a unit of this programme, P15 lost its own in this version when the designs named against it were found to measure no retained fraction of an installed gain, and P13 lost its own in the same version when the unit named against it was found to estimate a different quantity and to return an association where that proposition asks for a matched comparison. Two carry no confidence statement because the author has no view worth recording, and that is a refusal rather than an omission. Seven more carried no confidence at all in an earlier version of this document, which was a defect against the confidence field and was corrected rather than passed over, and were then held at a band across two rungs; the author has since stated each of the seven, five as a single rung, one across two rungs and one as a sentence of his own, so nothing in this document now waits on him for a confidence and the two refusals must not be read as those seven. Nineteen of the twenty-two take no other proposition with them, the exceptions being P1, P4 and P16, which is the point of stating them severably: this framework is built to lose parts and keep going. A reader who wants to end the conversion framework entirely should attack P1; a reader who wants to end the value two should attack P17, which is cheaper and which this document names as the weaker target on purpose. The invitation is asymmetric and deliberate: a refuted P17 removes the equality at one half, a supported P17 establishes one component of the condition set and never the value. Neither outcome is a measurement of temporal accumulation, which no registered instrument reaches.

P14 joins the no-instrument set in successor version 1.103, through the qualification in its entry and with its instrument line unchanged, because the design that line names estimates a different contrast. That correction adds no evidence, changes no confidence and scores no outcome.

SEVERABILITY AND SURVIVABILITY, AT THE LEVEL OF THE WHOLE DOCUMENT.

WHAT SURVIVES IF THIS UNIT FAILS. What this registration contains is a list of dated claims that fail separately. It offers no result of its own in support of any proposition here, so it has none to lose. Were every prediction here refuted, the dated record would still be exactly what it is: the 8 December 2024 manuscript, the 2 January 2026 print edition and the 2026 registrations would go on saying what was claimed and when, and the programme's measurements would go on being measurements. Refuting a prediction does not unmake the date it was made on. What would fall is the framework's claim to predict rather than describe.

WHAT SURVIVES AT THE LEVEL OF THE FIELD, AND WHAT DOES NOT, kept separate from the propositions because they fail independently. The measurement framework of Recursive Dynamics, the field the author founded, is the claim that systems which revise the machinery of their own improvement can be described by a small set of measured exponents: how capability scales with revision depth, how corrective service and newly generated burden scale along the same path, and what the balance between those two implies for stability. That framework is a proposal about what to measure. It survives the loss of the ceiling relation, the loss of the value two, and the loss of any particular exponent, because a framework is refuted by showing the quantities cannot be measured

or do not behave as quantities, not by a number coming out differently. What would end it is the balance exponent proving unidentifiable in principle, or the exponents proving unstable under the registered admissibility conditions in every domain tried. Neither has been shown, and neither is claimed here to be unlikely: the field's framework is a conjecture at the same standing as the laws, and it is named here so that a reader can see which of the two is being tested by which proposition.

THE ENGINEERING HYPOTHESIS IS SEPARABLE FROM BOTH. The protocol this programme designs, placing correction inside the improvement loop rather than outside it, does not follow from any law registered here and is not entailed by the framework. It is a hypothesis about a depth interaction, it is carried jointly by P12, P13, P15 and P21, with P7 its premise and P19 its motivation, and it can fail while every law here stands, or stand while they fall. It is registered as an engineering hypothesis under prospective test and is never presented as a consequence of a measured law.

SUCCESSOR QUALIFICATION. The method does not become a logical consequence of a law through the registration of Laws IV and V. The earlier separability clause refers to Laws I, II and III. Three of the Protocol's four carriers, P12, P15 and P21, are also the propositions of Laws IV and V, so a verdict on any of them bears on its law, under the verdict rules stated above, and on the method's engineering case by the share its own entry names.

WHAT A HOSTILE READER WILL SAY, PUT IN ITS STRONGEST FORM AND ANSWERED, because an objection a document does not state is one it cannot be trusted to have considered. Each answer below is a statement about this document's method and is checkable from this document. None of them makes any prediction stronger, and none of them is offered as evidence for anything.

FIRST: YOU INVENTED A LOCAL NOTATION BECAUSE YOUR OWN TERMINOLOGY WAS NOT STABLE. True, and the instability was found by this programme, before any confirmatory measurement, and is recorded with the date it was found. A collision between two estimands discovered after publication is a retraction; discovered before data it is a control. The register names the readings the bare letter has carried, the conversion between the two axes is printed here, and no exponent may be substituted for another. The question worth asking of any programme is not whether its notation ever collided, but whether the collision was found by the author or by the reader.

SECOND: THE CEILING IS CONDITIONAL ON AN ASSUMPTION YOU ADMIT IS NOT IDENTIFIABLE FROM AN ORDINARY OBSERVED PATH. True, and it is registered that way deliberately. An unconditional ceiling would be unfalsifiable decoration: it would survive every result because nothing would be specified that could contradict it. The conditions name the design that could refute it, crossed or off-path variation, and this document forbids scoring the relation without it. A prediction that says in advance which experiment could kill it is worth more than one that cannot be killed.

THIRD: NEITHER A FAVOURABLE NOR AN UNFAVOURABLE P17 RESULT ESTABLISHES THE VALUE TWO. Half true, and the half that is false is registered here rather than left for a reader to find. Refuting P17 removes the equality at one half on which the value rests. Supporting it establishes one component of the condition set and never the value. Both halves are printed in P17's

entry and beside the invitation to attack it. A proxy labelled as a proxy before it runs is not a substitution.

FOURTH: P6 CAN NEVER BE CONFIRMED, SO IT IS A CLAIM THAT SURVIVES EVERY NULL RESULT. True, and it is the one proposition here that can never pay its author anything. Its registered outcomes are refuted, no counterexample found among N, and inconclusive; the absence of a counterexample is never marked as support, and the scoring sheet says so on its own line. A universal claim its own programme cannot confirm is a standing invitation to be killed by a single counterexample from anyone. The asymmetry runs against the author.

FIFTH: YOU EXPECT A RESULT THAT PARTIALLY REFUTES YOUR OWN P19. True, and it is stated before any measurement. The entry records the cell the author expects, records that it refutes the wording registered here, and records that it is never reported as support. Registering a proposition one expects to lose is the thing registration exists for.

SIXTH: THIS IS TOO LONG AND IS TOO MANY THINGS AT ONCE. The length is the audit trail and the count is the severability. Twenty-two propositions that fall separately are not one claim: the dependency map states what each takes with it, the scoring sheet gives every line its own verdict, and a rule here forbids drawing any programme-level conclusion by counting supported propositions. A reader who wants the short version is told below exactly where it is.

SEVENTH: THE WHOLE SET COULD HOLD AND BEAR ON NO SAFETY OUTCOME. True, and the answer rests on the concession rather than on a rebuttal. The laws registered here are scaling relations between capability, the correction that services it and the drift it must overtake, and a relation of that kind holds or fails whatever a system's values are, so a supported set would establish scaling and stability relations and would establish nothing about misalignment or loss of control. That limit is not conceded here for the first time. The licence rules above state in the restrictive form what a supported P1, P4, P5, P16, P20, P21 and P22 would license, and refuse the safety reading wherever it is available; the whole set is registered as licensing no claim that a safe self-improving system can be built; no rung of the evidence ladder licenses the claim that alignment is solved or that any deployed system is safe; and a supported P4 is registered as a scaling condition and never a safety certificate. The two persistence registers carry the method rather than the laws and hold three propositions between them, which is the asymmetry named in the background. An answer written to make this objection go away would have to claim exactly what those licences forbid.

SUCCESSOR QUALIFICATION. The two persistence registers still carry the method and now also carry Laws IV and V. They carry no part of Laws I, II and III. This naming does not enlarge any instrument's endpoint or license a mechanism claim.

HOW TO AUDIT THIS DOCUMENT WITHOUT READING ALL OF IT, said plainly because a document nobody can check is not open to correction. Four surfaces decide how much of this can currently be tested at all, and they are named here rather than left to be inferred: the compressed list of the twenty-two; the scoring sheet, which carries every proposition's scope, confidence, instrument status and what it takes with it on one page; the propositions whose instruments do not decide them, which are the five with none at all, P12, P13, P14, P15 and P22, together with those blocked on a rule that is not yet settled; and the two confidences the author declined to state. If

those four surfaces disagree with the proposition bodies, the bodies govern and the disagreement is a defect to be reported. Nothing else in this document can change a verdict.

WHAT FALLS WITH THE UNIT AS A WHOLE. Only P1 can end the central hypothesis, and the dependency map says so.

WHAT THAT DOES NOT MEAN IS SET DOWN HERE BECAUSE THE WORD PROGRAMME IS USED TWO WAYS IN THIS FIELD AND THIS DOCUMENT ONCE USED BOTH. The ARC Programme is the body of dated work, and a refutation does not unmake a corpus: the papers keep their dates, the registrations keep their records, Recursive Dynamics keeps its proposal about which quantities are worth measuring, and the severable correction, evaluation and engineering propositions keep their instruments and their verdicts. What P1 ends is Law I, the integrated recursive-conversion hypothesis, and P2. Where any other passage in this document reads as though more than that were at stake, the narrower reading here is the one a scorer applies. Everything else comes apart cleanly, one piece at a time. Refute the ceiling value at P3 and the number moves. Refute the independence premise at P6 and the route to one half is gone. Refute the derivation claim at P5 and the exponent becomes a parameter that was fitted rather than predicted. Refute the corrector-class premise at P7 and the reason for moving correction at all goes; refute all four of the propositions that carry the protocol, P12, P13, P15 and P21, and its engineering recommendation goes with them. Refute the scoring claims at P10 and P14 and certain measurements lose their standing. Refute the embedding claim at P15 and an argument about where correction has to live goes. Not one of those takes the conversion relation down with it, and where a reader wants the authoritative answer to what any single proposition does take, it is the takes-with-it column and not this paragraph.

WHAT THIS UNIT DOES NOT TEST. Nothing. It measures nothing and tests nothing; every test must ultimately run under a separately registered instrument named beside its prediction, and at this date not one of those instruments is registered: they are drafts, blocked drafts or absent, exactly as each entry states, and the fixed-in-advance admissibility conditions decide what may count. A reader who finds a measurement cited against this registration should check it against those conditions before checking it against the prediction.

Update Criteria

SUCCESSOR RECORD NOTE. The following references to lodged version 1.102 files describe the frozen predecessor and remain historical references. Where a paragraph below calls version 1.102 the version of this registration, it speaks of the predecessor; this text is version 1.103. This text claims no PDF, manifest, public identifier or exposure certification of its own; the registry record carries its identifier, and the author's personal acts are recorded separately.

This registration is amended, never withdrawn, and the mechanics are stated correctly here rather than by an assumption about what withdrawal does. A permitted update identifies the field changed and the reason, and preserves the previously registered record beside it. Files attached to a submitted registration are immutable: a revised reader copy is lodged in the originating project's storage and is to be published at a persistent identifier the author assigns and links from this registration's resources, rather than replacing the attached one. Withdrawal is not an amendment. It

is irreversible and removes the substantive content, while the record retains its core metadata, including the creation, registration and withdrawal dates and the persistent identifier; an earlier wording here said a withdrawal loses the date, which is wrong and is corrected. The reason for preferring amendment is that the substance survives, not that the date would otherwise be lost.

WHAT THE REGISTRY ITSELF IS DOING, checked at its own notices and recorded here because it changes where the supporting material should live. The registry has announced that from 16 November 2026 no new projects or child components can be created on it, and that from 19 February 2027 existing projects become read-only, while registrations and preprints continue to operate as before and existing links and persistent identifiers continue to resolve. This registration is therefore built to be self-contained: its symbols travel with it in the extract above, its reasoning does not depend on a reader following a link into an editable project, and anything it relies on outside itself is cited by a persistent identifier rather than by a location that may become read-only. The development record, the reader copy, the rendered protocol and the checksum manifest of this version stand on those terms: each is lodged in the originating project's storage under its own filename, bound by the SHA-256 the manifest carries, the four being lodged there at the close of this version and before this text is submitted as an update, and each is to be published at a persistent identifier the author assigns and links from this registration's resources, that publication being a commitment rather than a description. Before any submission or update of this registration each field is read back from the draft and compared against this text, because a check that compares this text with itself cannot see a truncation introduced between the two, and the comparison is recorded rather than assumed.

WHICH FILE IS THE REGISTRATION, SAID PLAINLY BECAUSE THE FILE LIST IS LONG. The text of these fields is the registration. The file set that stands beside it, lodged in the originating project's storage and to be published at a persistent identifier the author assigns and links from this registration's resources, holds every earlier draft of that text, because this programme keeps its drafts rather than presenting a clean sheet, and the great majority of those files are therefore superseded by construction. The operative artefacts are the protocol file and the checksum manifest carrying the version stated at the head of this document, and that manifest lists every other file in the set with its role and its hash and marks each earlier protocol superseded. It cannot list itself, so its own hash is to be published with it at the persistent identifier this field commits to, and a reader verifying the set takes that one hash from there and every other hash from the manifest. The reader copy is verified against these fields by reading it against them, because the fields are the registration, and every other artefact is verified against the manifest's hashes wherever the set is published.

At submission the rule is replaced by the thing itself, and this version states what it does rather than promising more than it delivers: this text names the development record and the checksum manifest by literal filename, and the manifest carries the checksum of the reader copy, of the rendered protocol, of the development record and of the script that builds the render, as it carries the checksum of every other file in the set but its own. This text prints one checksum, the sixteen-character opening of the SHA-256 of the operational-definitions register the symbols above are taken from, which is not a file in this set, and it prints no checksum of any artefact of this version,

for the reason this field gives twice below: a checksum for a file rebuilt from this text cannot be taken before that file is finished, and a hash printed here for the development record could not describe the record a reader will open. It names no commit of the source it was itself built from, because the citable authority is the release named in the manifest with its checksums rather than a repository state, which the paragraph below on the approved release fixes. No later file joins that set by being numbered higher. The rendered protocol and its source are different representations of the same content and are not expected to share a hash; what is verified is content correspondence, with a separate checksum for each. Correspondence is established by reading the rendered file through, every page of it, against the source before it is lodged and published, because the failures a render introduces are exactly the ones a checksum cannot see: a table clipped at a page boundary, a blank page where a section ended, a substituted character standing in for the one written. The read is repeated at every render, since a file read at one version says nothing about the next.

A CHECKSUM THIS REGISTRATION PRINTS FOR A FILE THAT IS ITSELF REBUILT FROM THIS REGISTRATION CANNOT BE TAKEN BEFORE THAT FILE IS FINISHED, AND THE ORDER OF WORK IS FIXED HERE RATHER THAN LEFT TO BE REMEMBERED. The reader copy and its render are rebuilt from this text, the development record then closes with the entry reporting the page-by-page read of that render, and the manifest is made last, so no checksum stated anywhere in the set has to be taken before the file it names is finished. This field therefore names the development record by filename and leaves its hash to the manifest, because a hash printed in this text would have to be taken before the render this text builds and before the record's own entry about that render. No file in this set states the checksum of a file built after it: the manifest, lodged as theory-predictions-sha256-manifest-v1.102.txt and to be published at a persistent identifier the author assigns and links from this registration's resources, is the one place where every checksum is taken at the end and together, which is why it is the last artefact made and why it governs where two surfaces disagree about a hash. The history of the superseded selection rule is recorded in the development record.

AND THE AUTHORITY IS THE APPROVED RELEASE, NOT THE REPOSITORY HEAD. A repository branch moves, and anything that moves cannot be the object a scientific commitment is made about. The citable authority is the release named in the manifest with its checksums; later work in the same repository does not rewrite it, reopen it, or become it by being newer. A reader who finds the repository ahead of the release is looking at work in progress, and a reader who finds it behind is looking at a branch that has not merged, and neither changes what was registered. Those are the protocol and manifest whose filenames carry the version number stated at the head of this document.

THE POINTER TO THE OPERATIVE FILES IS WRITTEN AS A RULE RATHER THAN A NUMBER SO THAT IT CANNOT GO STALE, and the development record holds the history of the pointer that went stale.

THE SAME CORRECTION HAS NOW BEEN MADE IN THE BODY. Two passages above dated a superseded wording by the draft version number it was written at, and since every version of this text before 1.100 was an unlodged draft that no reader met, and version 1.101 was prepared as a draft and never submitted, so that no reader met that one either, a reader had no file to resolve

those numbers against and no sentence in this field covering them. Each of those passages names the earlier wording and its correction instead, and the version numbers are carried by the development record, where the chain is complete. No superseded draft states the position of this registration, several state positions it has since corrected, and the development record of this version, named in this field and checksummed in the manifest lodged beside it and to be published at a persistent identifier the author assigns and links from this registration's resources, records what changed and why in each case. A reader who quotes an earlier file as this registration's claim has quoted a draft this document itself withdrew.

WHY THE DEFINITIONS REGISTER IS CITED AND NOT REPRODUCED, since a reader may reasonably ask. These fields are self-sufficient for scoring: every quantity any of the twenty-two verdicts turns on is defined here in words, with its axis, its admissible range and the instrument that would measure it, and that was checked entry by entry against the register rather than assumed. Of the register's thirty-eight symbols, twenty-two appear here under the same names, and the load-bearing remainder appear here in prose: the balance exponent as the difference between the path elasticity of correction service and that of burden, the four-lever frontier, both direct depth-dependences, the boundary as a zero crossing of correction pressure, and the cost ratio that the allocation account turns on. What the register adds beyond these fields is its equation numbering, the derivation scaffolding no scorer needs, such as the constants of the saturating form, the two-point estimator, the combined test statistic and the standardised effect size, and its record of the symbol collisions this programme has suffered. It is cited by version, date and SHA-256, with the pinned version governing, rather than reproduced, for two reasons that are the same reason. A copy pasted here would be a third copy of one content, and the failure this document has hit most often is a compressed duplicate drifting from the entry it duplicates. And the register carries a defect this registration has already declared, its bare letter still on the checker axis; reproducing it would freeze that defect inside this document, where citation by hash instead lets the corrected successor be cited by amendment while the frozen record still shows exactly what was pinned on the day. Where the two disagree on an axis, this document governs, and that rule is stated above rather than left to precedence. Where a supporting resource can only be reached through such a project, that is recorded as a limitation of the citation rather than left for a reader to discover.

WHEN A PREDICTION IS REFUTED. It is RECORDED AS REFUTED IN THE SEPARATELY DATED STATUS RECORD, or through the registry's transparent update to the registered responses where that is the appropriate route, with the date, the measurement that refuted it and its identifier. An earlier wording said marked refuted in place, which reads as an edit to the original and is not what happens: the original submitted version and its attachments are preserved unchanged. It is not deleted, softened or reworded. The predictions it names as falling with it are marked as fallen in the same amendment, whether or not they have been tested separately, because that dependency was registered in advance and honouring it is the only thing that makes stating it worth anything.

WHEN A PREDICTION IS SUPPORTED. It is marked supported with the same detail, and the registration states explicitly whether the supporting measurement met every admissibility condition applicable to it. A measurement that supports a prediction while failing an admissibility condition is recorded as inadmissible support, which counts for nothing here.

WHEN AN INSTRUMENT IS BUILT FOR P12, P13, P14, P15 OR P22. The trigger therefore covers P12, P13, P14, P15 and P22, and the count of five agrees with the compressed list and the scoring sheet, which is what naming the propositions in both places was for. Every movement of names into and out of that set is recorded in the development record; P14's entry into it is recorded by the qualification this successor places in P14's entry and in this successor's change table. The instrument is registered separately before it runs, and its identifier is carried by the transparent-changes record, or by a registered successor where the change is substantive, never by editing this text once it is frozen. Building an instrument and then reporting that it confirmed a prediction registered here, without a separate preregistration of the instrument, would defeat the purpose of both documents.

WHEN THE FRAMEWORK CHANGES. If a derivation is corrected, the correction is recorded here with what it changed and why, in the same form as the three reversals disclosed in the background. A theory-level registration whose history shows corrections is more credible than one that never needed any; a theory-level registration that quietly acquires new predictions is worth nothing.

WHAT WOULD END THE PROGRAMME RATHER THAN AMEND IT. Refutation of P1, which is the frame. If capability does not scale as a power of depth in the tested class, the author commits to saying so as the headline finding rather than rescoping the claim until it survives.

WHAT NO AMENDMENT MAY DO. No amendment may add a prediction and present it as having been registered on this date, narrow a scope condition after a measurement has been seen, or convert an inconclusive outcome into support. Each of those is the specific manoeuvre this document exists to make impossible, and each would be visible in the version history.

HOW AN AMENDMENT IS ACTUALLY MADE, because a protocol without an instrument is a wish. This registration was amended through the version history while it remained a draft. Once submitted it is frozen and is not edited in place. A post-submission change is carried by the registry's own transparent update to the registered responses, which identifies the fields changed and the reason and preserves the previous responses beside the new, or by a separately versioned transparent-changes record, and, where the change is substantive, by a newly registered successor that links to the original and states exactly what changed. The original registered record is never replaced or rewritten. An earlier wording here promised an amendment facility of the registration itself, and a later one said that a frozen registration does not provide one.

WHAT THIS SUCCESSOR CHANGES, STATED EXACTLY AS THE PRECEDING PARAGRAPH REQUIRES. Version 1.103 succeeds version 1.102, which stands unchanged beside it, as does the first registered version. Each line below is a line of the plain-text copy of version 1.102 published on the programme's site (SHA-256 dd362aea9c61c152f0d335079ab0bb7d1548ab37fe6459c6a1f41f381195115a), with what supersedes or qualifies it in this text; every line not listed stands as registered.

• Lines 76, 88, 682 and 706: the naming of the laws. Version 1.102 names three; at version 1.103 Laws IV and V are registered beside Laws I, II and III, whose statements stand exactly as registered. Line 76 is followed in this text by a qualification dating the idea each added law formalises from the author's earliest record, as the idea and not the law.

• Line 696: followed in this text by how a law takes its word, by the verdict rule of Law III, whose statement stands as registered, and by the statements, kill conditions and verdict rules of Laws IV and V, none of which version 1.102 carries.

• Lines 122 and 126: ARC-4 and ARC-5, which now test Laws IV and V, with no proposition entering or leaving either register.

• Lines 84, 106, 700, 1354 and 1370: the method's separability and the untouched laws, read as Laws I, II and III, since the method's carriers P12, P15 and P21 are also the propositions of Laws IV and V.

• Lines 170, 382, 704 and 796: the propositions that bear on the laws and the rule on which of them a reader weighs. Under version 1.103 fifteen propositions carry registered laws and seven carry none.

• Line 660: the naming of Laws II and III, answered only as far as that paragraph itself shows; the exact relation between the two laws is left to a later successor.

• Lines 1068, 1110, 1112, 1262 and 1376: what a refuted P12, P15 or P21 costs, read with the verdict rules registered in the Predictions field.

• Lines 124, 1082 and 1216: statements that a proposition in no register touches no law. They stand, and are read of Laws I to V.

• Lines 384, 386, 388, 464, 480, 492, 758 and 1344: the scoring gate, which stands as registered for every margin with a judged term, read with the counted route this successor extends to the margins of P21 and P17.

• Line 1250: P21's reason for registration, followed in this text by a qualification that names the protocol it speaks of as the Eden Protocol, as line 796 now does, with the entry's registered words unchanged.

• Lines 1252, 1254, 1256, 1258 and 1268: P21's refuter, its deciding unit and its instrument line, each unchanged, read with the directional reading, the two verdicts and the counted instrument this successor adds beside them.

• Lines 1140, 1148, 1152, 1154 and 1164: P17's population, its refuter, its mapping of the unit's four labels onto marks, its rule for a panel that returns more than one label and its instrument line, each unchanged, read with the reading of what a corrector is, the reading of when a correlated pair is a refutation and the counted instrument this successor adds beside them in P17's entry; line 516, the general rule that replication gates refutation, read with the second of those readings; lines 486 and 742, where the instrument is to state what independence is over and where the Hypotheses field states P17's panel, read with the first; and line 1156, the attribution reading printed beside any refutation, read with the counted route as the paragraph after line 480 states.

• Line 1098, with the inventories at lines 16, 64, 68, 198, 794, 1320, 1330, 1346, 1372 and 1406: P14's instrument status, corrected to NONE by a qualification placed in its entry after line 1096,

with its instrument line unchanged. Lines 52 and 190, which count what the instrument lines name, keep their registered words, each with a clause added on P14's position.

• Lines 1458 and 1464: the References entries for the book and for the Operational Definitions, each followed in this text by a link, the book's official record online and the register's persistent identifier, minted on 4 October 2026, with the entries' registered words unchanged.

• Line 1460: the References entry for the statement paper, followed in this text by the version of that paper this successor reads and quotes, 7.3, with the entry's registered words unchanged.

• Lines 1468 and 1509: the commitment to engage the external work before scoring, discharged for P21 and P17 alone by the two entries this successor adds to the References field; line 1468's statement that none of the anchors named there is engaged beyond being named is read as of version 1.102, since those entries engage six of them: Huang and colleagues (2023) in P21's, and Knight and Leveson (1986), Eckhardt and Lee (1985), Littlewood and Miller (1989), Kim and colleagues (2025) and Goel and colleagues (2025) in P17's.

• Lines 82, 1352 and 1462: Recursive Dynamics, named as the field the author founded, with its founding paper cited by the title its record carries. Line 82 is followed in this text by a qualification dating the term from the author's earliest record.

• Line 202: followed in this text by a dated status recording the public OSF component of 5 October 2026 that holds the record of 8 December 2024 and the manuscript of 30 April 2025; line 70's words, a private statement of December 2024, stand and are read as of version 1.102.

• Line 490: followed in this text by a qualification reading the blinding finding at the width Paper IV.d gives it.

• Line 1488: followed in this text by a qualification naming the prior work nearest to the direction this successor gives Law V's build-order limb, with what each states.

• Line 1511: followed in this text by a qualification lifting the testimony downgrade for the three assertions that now carry identifiers, and for them alone.

• Line 1414: followed by this list.

• Lines 1, 3, 5, 7, 9, 11, 22, 24, 32, 402, 982, 1380, 1466 and the closing generation note: the identity, dates and dated status of this text. Lines 114, 162, 322, 484, 542, 686, 694, 700, 1386, 1390, 1398, 1416 and 1426 are retained and say this version or until this version, as do the changed lines 198 and 1346, and lines 404, 406, 984, 1313 and 1378 are retained and state a registry or instrument status in the present tense, as do the clauses of lines 52 and 190 that call the instruments unbuilt and unrun: each is read as of version 1.102, as the status paragraph after line 32 says of every retained instrument-status sentence and the historical snapshot note before line 402 says of lines 404 and 406; lines 984 and 1313 speak of P7's instrument, the corrector-class audit, whose later registration and run as Unit 06 the dated statuses after lines 32, 402, 982 and 1466 record.

Three statements stay true in this successor and change only when a later successor registers the proposition of Law IV's comparison limb: line 122's statement that ARC-4 holds P15 alone, line 124's

partition of the twenty-two among the registers, and the sentence of line 1112 that the comparison between placements is P21's. Nothing listed here changes a proposition's number, words, refuter, instrument line, confidence, kind, scope, register assignment or date, the statements of Laws I to III and the ARC Bound, the dependency map (lines 358 to 374), the licence rules and the evidence ladder (lines 756 to 782), the four marks (line 1302) or the amendment rules (lines 1402 to 1420), whose list of propositions without an instrument, at line 1406, gains P14.

WHAT THE REGISTRY DOES PROVIDE AND WHAT IT DOES NOT PERMIT. The registry does provide a transparent update process for the registered responses, with the original version preserved and the change visible. What it does not permit is replacing or updating the files attached to a submitted registration. These fields are the author's, and the way he prepares them produces draft registrations rather than amendments to a submitted one, which is a limitation of this programme's own arrangements and not of the registry. A correction of the registered text itself, such as this version, therefore goes through the registry's own update, which keeps the earlier responses visible beside the new. This programme chooses the stricter route for later verdicts and for later work, which go to a separately dated status record or to a registered successor rather than into the registered text, and the original is never rewritten. That gap is recorded here rather than discovered at the moment the first amendment is needed. It is the same class of gap this document reports against P5 and P12, and it would be inconsistent to disclose theirs and conceal its own.

THE CODE-DEFECT PROTOCOL. It is a correction procedure, so it belongs with the amendment rules. The complete statement is kept below and the partial duplicate is dropped. Nothing in the commitment itself changes.

A defect is a disagreement between the implementation and the specification. If, after registration, the implementation is found to compute something other than what the analysis plan specifies, the specification governs and the implementation is corrected. Every such correction is recorded in an append-only defect log carrying the date, the defect, the diff, and the reason it is a defect against this text rather than a preference about the output. The log is published with the results whether or not any correction proves necessary. A preference is not a defect. Disliking the result, wanting a different estimator, or finding a more powerful method is expressly outside this clause. Those are exploratory analyses, are labelled as such, and never displace the registered one. After outcomes are seen the door closes on changes to the specification, and it does not close on conformance to it. Two cases are separated. Where the code did not execute the frozen specification, the erroneous output is preserved and published, the code is corrected to the specification, it is re-run, and the conformant result is the confirmatory one, because the registered commitment is to the specification rather than to a program that contradicts it. Where the specification itself is altered after any registered outcome has been computed on collected data, the altered analysis is exploratory or a sensitivity analysis and never confirmatory. Both numbers are published in either case, and the defect log records which of the two applied. A deciding unit retains the pre-blinding artefacts, the condition assignment and the laundering record, so that a blinding defect found after the fact is corrected and re-run under the clause above that governs an implementation defect.

THE DEVELOPMENT RECORD IS LODGED AND IS TO BE PUBLISHED RATHER THAN CARRIED IN THIS FIELD, AND THE REASON IS THE SAME REASON THIS DOCUMENT SEPARATES

REGISTRATION STATUS FROM EVIDENTIAL STATUS. Every version of this text from 1.0 of 24 August 2026 to VERSION 1.102 OF 10 SEPTEMBER 2026, which is the version of this registration and is stated here so that a field read on its own identifies which version it belongs to, existed as a draft. Every version before 1.100 was an unlodged draft that no reader met, and version 1.101 was prepared as a draft and never submitted, so no reader met that one either, and a log of those changes is a working record rather than a correction to a public record. This version corrects wording carried by version 1.100, and as this update is read the first version stands beside it, so that correction is carried by the update's justification and by the fields it changes rather than by this record, which remains what it has been throughout, the dated account of how the text was written. Until version 1.88 it was carried here, where it had grown to a third of the whole registration and was longer than the twenty-two propositions combined, which put a changelog of unregistered drafts in front of the predictions this document exists to fix. The development record is lodged in full and unedited, in the originating project's storage, as theory-predictions-development-record-v1.102.md, and is to be published at a persistent identifier the author assigns and links from this registration's resources; its sha256 is carried by the checksum manifest, which is regenerated last, carries the hash of every file in the set but its own, and is lodged beside it and to be published at a persistent identifier the author assigns and links from this registration's resources.

A HASH FOR THIS RECORD PRINTED IN THIS FIELD COULD NOT DESCRIBE THE RECORD A READER WILL OPEN, AND THE REASON IS AN ORDER OF WORK RATHER THAN AN OVERSIGHT. Hashing the record into this text would fix its number before the render this text builds, and the entry that closes the record is the one reporting the read of that render, so a number printed here would always be taken before the entry that completes the file it names. The manifest is made after every other artefact, is lodged with the set in the originating project's storage and is to be published at a persistent identifier the author assigns and links from this registration's resources, so it is the one surface on which this record's checksum describes the record a reader will open, reachable in that storage once the set is lodged and at that identifier once it is assigned. Nothing was removed from it and nothing was rewritten in it. A reader who wants to know what this text said at any earlier hour, and why it stopped saying it, has the whole of it.

WHAT THAT RECORD IS OFFERED AS, AND WHAT IT IS NOT. It is offered as evidence that defects were found and fixed against the author's own interest before any reader required it, that several reported defects were checked and not adopted with the reason given, and that two corrections were themselves wrong and were withdrawn rather than quietly dropped. It is not offered as a record of published corrections: this version makes exactly one, to wording carried by version 1.100, and that correction is carried by the update's justification and by the fields it changes rather than by this record, which stays the dated working record of the drafts. It carries no proposition, no verdict, no scope and no confidence, and nothing in it can be cited as a commitment of this registration; the fields above are the registration and the record is a companion to them.

WHERE THE SUPERSEDED WORDINGS THEMSELVES ARE NAMED. The propositions and rules above still name, in the sentence that governs, any reading this document has withdrawn, because a scorer meeting a rule needs to know which reading is unavailable to them. That is a forward

instruction rather than history, it stays where a scorer will meet it, and it is checked mechanically on two limbs: a withdrawn wording may appear in the fields above only inside the sentence that disclaims it, and a figure of the row printed in the background may appear in the fields above only with the estimator it belongs to.

Conflicts of Interest

The author has a direct interest in these predictions being correct. The framework is the subject of a book in print and of a research programme the author is seeking to fund, and a refutation of P1 would end both. That interest is why this registration exists in this form: severable predictions with refutation conditions fixed in advance are harder to bend than a narrative, and the three reversals are disclosed in the background rather than left for a reader to discover.

No external funding supports this work. There is no institutional affiliation, no grant, and no external body with a stake in the outcome. The author is not employed in the field.

Michael Darius Eastwood conceived and directs this research programme and is the author of this work. Across the programme, he has used more than six AI systems in parallel, under his own instructions, to stress-test his arguments, identify possible errors, and assist in preparing draft text from his own outlines. He determines what is adopted, revised or rejected and takes responsibility for the published content. These systems are tools, not authors.

One further disclosure, because it bears on how a reader should weigh the confidence levels above. The programme's own checks have repeatedly found errors in the programme's own materials, including, in the twenty-four hours before this registration was written, a retracted relation still displayed as live inside an analysis file, and an outcome measure defined two incompatible ways in one document. Those were found and corrected before registration. A reader should assume that the rate of such errors is not zero going forward, and the amendment protocol above is written on that assumption rather than in hope.

Acknowledgments

The three reversals disclosed in the background were each found by adversarial review rather than by the author's own satisfaction with the work: the exponent by a cross-architecture replication that failed, the ceiling relation by a directed interrogation of the derivation, and the self-acceleration expression by a reviewer following it through to its consequences. The programme's practice of keeping an append-only defect log, and of publishing that log with results whether or not it ends up empty, comes from those three occasions.

The community-developed template this registration follows is acknowledged in the template attribution section of the form.

The operational-definitions register, an extract of which is lodged in the originating project's storage and to be published at a persistent identifier the author assigns and links from this registration's resources, exists because four separate surfaces of this programme had been defining the same

symbols by hand, and on 22 August 2026 three of them asserted a ceiling relation that had been corrected six days earlier. One generated register replaced four hand-written ones.

References

Eastwood, M. D. (2026). Infinite Architects. Print edition, 2 January 2026, ISBN 978-1806056200. Section A.3, Testable Predictions, of Appendix A carries the depth-scaling prediction quoted in the background, dated by the print itself rather than on its own face; a separate Appendix F carries the book's wagers, three of them dated, the furthest to 2029.

SUCCESSOR QUALIFICATION. The book is published independently, with the author as its publisher, so its official record online is the publisher's page, https://www.michaeldariuseastwood.com/book/, which names this edition and its ISBN (read on 15 October 2026). This entry's registered words stand unchanged.

Eastwood, M. D. (2026). The ARC Theory: Statement Paper. Dated 1 September 2026, the date recorded in the alternative baselines above. DOI 10.17605/OSF.IO/GW5MX, the identifier given in the background above. Listed here because a reader looking for the citation looks in this field, and no claim of that paper is carried into this document.

SUCCESSOR QUALIFICATION. The version of that paper this successor reads is 7.3, revised 7 October 2026, at the same identifier. Its Table 1 rows for ARC-2 and ARC-4, the sentence of its section 7 on the persistence trials and the sentence of its note of 3 October 2026 are quoted in the Predictions field, where this successor narrows them before any data. No claim of that paper is registered here.

Eastwood, M. D. Founding Paper · Recursive Dynamics · The Science of Systems That Improve Themselves. DOI 10.17605/OSF.IO/HCPBU, the identifier given in the background above. No date or version for it is printed in this registration and none is supplied here, so it is cited by title and identifier alone. Listed here on the same terms, and no claim of that paper is carried into this document.

Eastwood, M. D. (2026). Operational Definitions of the ARC Programme, version 1.7.1, 3 September 2026. The register defining every symbol and equation used in this registration, with a plain-language reading and a measurement procedure for each, the symbol collisions this programme has suffered, and the retracted forms with the single reading under which each remains legitimate. An extract of it is lodged in the originating project's storage and is to be published at a persistent identifier the author assigns and links from this registration's resources, and the register itself is neither lodged nor to be published with it; the register's persistent identifier is recorded in this registration's history when minted.

SUCCESSOR QUALIFICATION. The register's persistent identifier was minted on 4 October 2026: https://doi.org/10.17605/OSF.IO/GVQ82, the OSF project Register · The ARC Theory · Operational Definitions. The versions it holds begin on 4 October 2026, after the version 1.7.1 named here. This entry's registered words stand unchanged.

HISTORICAL SNAPSHOT FROM VERSION 1.102, PUBLISHED 13 SEPTEMBER 2026. The retained paragraph below reports that earlier record; its present-tense status words do not assert the status at the date of successor version 1.103.

The studies named as instruments in the prediction list are separate draft registrations in this programme, each with its own analysis plan and design sensitivity. None is lodged with any registry and none yet carries a persistent identifier, which is said plainly here because this document polices that word everywhere else. Each identifier is carried by the transparent-changes record as it is minted, and never by editing this text once it is frozen. No result from any of them exists at the time of writing.

DATED STATUS, 3 OCTOBER 2026 AT 00:09:45 UTC. Unit 06, the corrector-class audit, was submitted on 1 October 2026 at 17:13:36 UTC under OSF identifier rv2b5. At the stated observation time OSF recorded public=false and pending_registration_approval=true: SUBMITTED, PENDING APPROVAL, not approved and public. All 19 archived files and all 23 text fields matched the deposited version; the six attachment fields also matched by file identity and hash. These are registration and text-conformance facts, not a study run, a proposition verdict or a certification of what the author knew. No study outcome is inferred.

DATED STATUS, 11 OCTOBER 2026 AT 19:01:59 UTC. Unit 06, the corrector-class audit, is registered under OSF identifier rv2b5 (DOI 10.17605/OSF.IO/RV2B5): at the stated observation time OSF recorded it public and no longer pending approval, and its public results record it as shown approved and public at 05:10:36 UTC on 4 October 2026. Its registered run took place on 6 October 2026. Run A completed 303 calls; Run B returned HTTP status 402 at its first call, which the registered failure policy makes final, so no row was counted. The study is reported as administratively incomplete under its registration, with no inferential verdict. The results are public at https://osf.io/4bphg/ (DOI 10.17605/OSF.IO/4BPHG), and the registration rv2b5 links to them as its data resource. These are registration and run facts as OSF's records and the unit's public results state them; no proposition verdict is inferred from them.

EXTERNAL ANCHORS NAMED AT THE TIME OF REGISTRATION, without commitment to full engagement. Each is the work a reviewer would name first against the prediction it is listed under; each was verified against its source record, and the arXiv entries below were re-checked against the arXiv metadata record, most recently on 4 September 2026, twenty-eight at that check against the twenty-nine now listed, with title, first author and first-version date matching the attribution made here, the one outside that check being the Schmidhuber entry, which was added after it, and the journal entries were checked against their publisher records on the same day; none is engaged here beyond being named, and the commitment below to engage the literature before scoring stands unchanged.

Against P4, P7, P13 to P15: Chen, Y.-H., Wen, J. and Kirchner, J. H. (2026). Automated Researchers Can Mitigate Well-characterized Alignment Failures. arXiv:2608.28945, first version 28 August 2026, third version 2 September 2026, title given as published, convergent apparatus and never confirmation.

Against P6: Knight, J. C. and Leveson, N. G. (1986). An experimental evaluation of the assumption of independence in multiversion programming. IEEE Transactions on Software Engineering SE-12(1), 96 to 109, doi 10.1109/TSE.1986.6312924; Eckhardt, D. E. and Lee, L. D. (1985). A theoretical basis for the analysis of multiversion software subject to coincident errors. IEEE Transactions on Software Engineering SE-11(12), 1511 to 1517, doi 10.1109/tse.1985.231895; Littlewood, B. and Miller, D. R. (1989). Conceptual modeling of coincident failures in multiversion software. IEEE Transactions on Software Engineering 15(12), 1596 to 1614, doi 10.1109/32.58771.

Against P7 and P14: Burns, C. et al. (2023). Weak-to-Strong Generalization: Eliciting Strong Capabilities With Weak Supervision. arXiv:2312.09390; Bowman, S. R. et al. (2022). Measuring Progress on Scalable Oversight for Large Language Models. arXiv:2211.03540.

Against P1 and P8, and named first because it holds the priority this document concedes: Sharma, A. and Chopra, P. (2025). The Sequential Edge: Inverse-Entropy Voting Beats Parallel Self-Consistency at Matched Compute. arXiv:2511.02309, 4 November 2025, prior publication of the matched-compute sequential advantage, never independent corroboration of it. Also against P1, and adverse to it: Wu, Z. et al. (2025). Is Depth All You Need? An Exploration of Iterative Reasoning in LLMs. arXiv:2502.10858, February 2025, which sets iterative depth against breadth and reports breadth matching or outperforming depth. Also against P1, and the most recent work pointing away from it: Gu, X. et al. (2026). Understanding Performance Gap Between Parallel and Sequential Sampling in Large Reasoning Models. arXiv:2604.05868, 7 April 2026, which compares the two across Qwen3, DeepSeek-R1 distilled models and Gemini 2.5 and reports parallel outperforming sequential, attributing the gap to reduced exploration rather than to the aggregator or to context length. Also against P1, and a result for the combination rather than for either arm: Bilal, A. et al. (2026). Refining Over Resampling: Test-Time Self-Correction for LLM Reasoning. arXiv:2608.05643, 6 August 2026, which samples independent attempts, refines each of them, and beats both wider sampling and the verifier-based and search baselines it tests. Also against P1, as the empirical basis of the breadth rival: Brown, B. et al. (2024). Large Language Monkeys: Scaling Inference Compute with Repeated Sampling. arXiv:2407.21787, first version 31 July 2024, which reports that coverage "scales with the number of samples over four orders of magnitude" and is "often log-linear and can be modelled with an exponentiated power law", which is the shape a depth-only measurement could be mistaken for. Also against P1, and named because it owns the term this proposition uses: Alabdulmohsin, I. and Zhai, X. (2025). Recursive Inference Scaling: A Winning Path to Scalable Inference in Language and Multimodal Systems. arXiv:2502.07503, first version 11 February 2025, which introduces recursive depth as a compute-matched inference-scaling recipe and derives data scaling laws it reports as improving both the asymptotic limit and the scaling exponents. Also against P1, as the nearest published exponent carried by a recurrence count: Schwethelm, K., Rueckert, D. and Kaissis, G. (2026). How Much Is One Recurrence Worth? Iso-Depth Scaling Laws for Looped Language Models. arXiv:2604.21106, first version 22 April 2026, measuring that exponent at 0.46, later than this programme's own paper and therefore not an antecedent to it. Also against P1, as the empirical basis of the saturating rival: Prairie, H. et al. (2026). Parcae: Scaling Laws For Stable Looped Language Models. arXiv:2604.12946, 14 April 2026, reporting test-time looping that scales "following a predictable, saturating exponential decay". Under the balance formulation recorded in the alternative baselines that result stops being only adverse and becomes

something this framework predicts about: a saturating curve has a falling local elasticity, so the framework says such a system is correctable and increasingly so. That is a prediction this framework could lose and it is not a confirmation of anything. It is checkable in principle on systems of the kind that paper studies, once correctability is measured on them; that paper did not measure it. Also against P1, on the allocation between widening and deepening: Inoue, Y. et al. (2025). Wider or Deeper? Scaling LLM Inference-Time Compute with Adaptive Branching Tree Search. arXiv:2503.04412, 6 March 2025. Also against P1, on recursive depth as computation rather than as artefact revision: Geiping, J. et al. (2025). Scaling up Test-Time Compute with Latent Reasoning: A Recurrent Depth Approach. arXiv:2502.05171, 7 February 2025. Also against P1, in another modality and one day before this programme's own paper: Jaiswal, S. et al. (2026). Iterative Refinement Improves Compositional Image Generation. arXiv:2601.15286, 21 January 2026. Also against P1, as a power law in a depth-like quantity on a different axis: Wang, T. et al. (2026). Can RL Teach Long-Horizon Reasoning to LLMs? Expressiveness Is Key. arXiv:2605.06638, 7 May 2026. Also against P1, as the closest published work in motivation to artefact-mediated recursion: Kim, Z. M. et al. (2026). Meta^n: Recursive Self-Improvement through Emergent Depth. arXiv:2608.24735, 25 August 2026. Also against P1, on the allocation of test-time compute, and earlier than both parties on the comparison this document makes: Snell, C. et al. (2024). Scaling LLM Test-Time Compute Optimally can be More Effective than Scaling Model Parameters. arXiv:2408.03314, 6 August 2024, which sets revisions in sequence against equally many parallel attempts, finds sequence narrowly ahead in its own words and calls the two complementary axes, four months before this programme's dated record, so no precedence on the direction or on the comparison is claimed here. Also against P1 and P5, as the self-improving agent lineage this programme does not claim to have originated: Robeyns, M. et al. (2025). A Self-Improving Coding Agent. arXiv:2504.15228, 21 April 2025; Zhang, J. et al. (2025). Darwin Godel Machine: Open-Ended Evolution of Self-Improving Agents. arXiv:2505.22954, 29 May 2025. Also against P1, as the published form of the broken-power rival named in its functional-form comparison: Caballero, E. et al. (2022). Broken Neural Scaling Laws. arXiv:2210.14891, first version 26 October 2022, which proposes a smoothly broken power law and fits it across a wide range of scaling behaviours. A short ladder cannot separate that form from a single power law, which is why it is carried as a rival rather than mentioned as a caveat.

Against P4 and P19, and the most important collision either audit returned: Engels, J. et al. (2025). Scaling Laws For Scalable Oversight. arXiv:2504.18530, 25 April 2025, which quantifies the probability of successful oversight as a function of overseer and overseen capability, fits scaling laws across four oversight games, and derives the optimal number of levels for nested oversight; its reported shape is piecewise linear with two plateaus, and its existence means no claim of first quantification of oversight scaling is available here.

Against P17 and P6, alongside the error-correlation work already named: Kim, E. et al. (2025). Correlated Errors in Large Language Models. arXiv:2506.07962, 9 June 2025, over 350 models, reporting agreement 60 per cent of the time on one leaderboard dataset when both models err, with correlation higher for larger and more accurate models even across distinct architectures and providers.

Against P15: Qi, X. et al. (2023). Fine-tuning Aligned Language Models Compromises Safety, Even When Users Do Not Intend To! arXiv:2310.03693, 5 October 2023.

Against P12: Ung, M. et al. (2024). Chained Tuning Leads to Biased Forgetting. arXiv:2412.16469, 21 December 2024.

SUCCESSOR QUALIFICATION. The direction this successor gives Law V's build-order limb, that a corrector formed before the capability it corrects has the larger correction-leverage exponent, has prior work nearer than the anchor above, named here with what each states and no more, each read at its arXiv record on 15 October 2026. Korbak, T. et al. (2023). Pretraining Language Models with Human Preferences. arXiv:2302.08582, first version 16 February 2023, published at ICML 2023 as its record states: they compare five objectives for pretraining with human feedback on three tasks and report that pretraining with human feedback "results in much better preference satisfaction than standard LM pretraining followed by finetuning with feedback", which they describe as "learning and then unlearning undesirable behavior", their best objective, conditional training, reducing the rate of undesirable content "by up to an order of magnitude" "both when generating without a prompt and with an adversarially-chosen prompt"; they measure preference satisfaction and rates of undesirable content, and not a correction-leverage exponent, a slope against the capability corrected, the registered maximal contrast at matched content or persistence under later capability training. Maini, P. et al. (2025). Safety Pretraining: Toward the Next Generation of Safe AI. arXiv:2504.16980, first version 23 April 2025: they propose "a data-centric pretraining framework that builds safety into the model from the start", hold that "Post-hoc alignment methods are brittle: once unsafe patterns are learned during pretraining, they are hard to remove", and report attack success rates on safety benchmarks falling from 38.8 to 8.4 per cent; they measure no exponent and no order contrast at matched content. That alignment pressure applied from the start of training can beat the same pressure applied afterwards is therefore prior on those endpoints, Korbak and colleagues' date preceding the author's record of 8 December 2024, and the limb claims no priority for it. What the limb registers that neither states is an exponent difference of at least the registered minimum effect, on the target axis, on the registered maximal contrast at matched content, as a law under test with its kill condition. The anchor above stands as named.

Against P17 and P22: Goel, S. et al. (2025). Great Models Think Alike and this Undermines AI Oversight. arXiv:2502.04313, first version 6 February 2025, on error correlation among capable models, which bears directly on the independence premise and points against it. Against the form of the programme rather than any one proposition, as the prior statement of the form: Maxwell, J. C. (1868). On Governors. Proceedings of the Royal Society of London, volume 16, pages 270 to 283, read 5 March 1868, which connects a model of the machinery, a criterion separating a decaying disturbance from a growing one, the onset of instability as the adjustment changes, and the practical remedies, in a single paper. Against the claim that self-improvement is not a subject existing disciplines can formalise, which this document does not make: Schmidhuber, J. (2003). Goedel Machines: Self-Referential Universal Problem Solvers Making Provably Optimal Self-Improvements. arXiv:cs/0309048, first version 25 September 2003, latest version 17 December 2006, which formalises a system that rewrites any part of its own code once it has proved the rewrite beneficial under its own axioms.

Against the scoring instrument and conditions A and E: Li, W. et al. (2026). Grading Scale Impact on LLM-as-a-Judge: Human-LLM Alignment Is Highest on 0-5 Grading Scale. arXiv:2601.03444, 6 January 2026, which reports that "the grading scale of 0-5 yields the strongest human-LLM alignment", a finding about the instrument this programme's scorer is built on.

Against P4 and P15: Huang, J. et al. (2023). Large Language Models Cannot Self-Correct Reasoning Yet. arXiv:2310.01798; Kumar, A. et al. (2024). Training Language Models to Self-Correct via Reinforcement Learning. arXiv:2409.12917.

Against P9, and named because that proposition concedes the dimensional form to this literature rather than claiming it: West, G. B., Brown, J. H. and Enquist, B. J. (1997). A General Model for the Origin of Allometric Scaling Laws in Biology. Science 276(5309), 122 to 126, doi 10.1126/science.276.5309.122; Banavar, J. R., Maritan, A. and Rinaldo, A. (1999). Size and form in efficient transportation networks. Nature 399(6732), 130 to 132, doi 10.1038/20144; Demetrius, L. (2003). Quantum statistics and allometric scaling of organisms. Physica A 322, 477 to 490, doi 10.1016/s0378-4371(03)00013-x; Demetrius, L. (2006). The origin of allometric scaling laws in biology. Journal of Theoretical Biology 243, 455 to 467, doi 10.1016/j.jtbi.2006.05.031. What P9 registers is the assignment of an effective dimension from how a domain composes its inputs and outputs, and the fact that the assignment precedes the curve. The relation between dimension and exponent is theirs.

Against P5 and the coupling relation, and named because it is the antecedent a reader from another field reaches for first: the form in which an exponent is one over one minus a coupling is standard in the economics of ideas, where it carries the long-run growth rate of knowledge rather than a growth exponent of capability. Romer, P. M. (1990). Endogenous Technological Change. Journal of Political Economy 98(5, Part 2), S71 to S102, doi 10.1086/261725; Jones, C. I. (1995). R and D-Based Models of Economic Growth. Journal of Political Economy 103(4), 759 to 784, doi 10.1086/262002, in which the semi-endogenous form makes long-run growth turn on exactly such a ratio. This registration claims no priority over that form, and a reader who recognises it has recognised something real. What is registered here is narrower: that the coupling is measured on records disjoint from the growth curve and the implied exponent sealed before the validation trajectories are revealed, which is a different act from writing the form down. Adverse to the same family, and named for that reason rather than despite it: Bloom, N., Jones, C. I., Van Reenen, J. and Webb, M. (2020). Are Ideas Getting Harder to Find? American Economic Review 110(4), 1104 to 1144, doi 10.1257/aer.20180338, which measures research productivity as falling sharply across several fields. That is an antecedent pointing against sustained compounding in the one literature that has measured it longest.

Against the framework as a whole: Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press. It states the control problem, the question of how a less capable party keeps oversight of a more capable one, and that question is the standing problem this programme's laws try to put a number on.

Miller, A. C. (2024). Registering Theory-Based Predictions in Political Science. PS: Political Science and Politics, Cambridge University Press, published online 25 October 2024, open access under CC-

BY. The template this registration follows, acknowledged as the form requires. https://www.cambridge.org/core/journals/ps-political-science-and-politics/article/registering-theorybased-predictions-in-political-science/9D022F2678C5BBBBE0ABCB704E3EAE0A

WHAT THIS REFERENCE LIST DOES NOT YET CARRY, stated because the gap is visible anyway. The development record holds the sentence withdrawn from here as stale. What remains true is the substantive gap: the external work is listed rather than engaged, and for most propositions this document does not record the closest prior claim, the closest existing apparatus, the strongest rival explanation, or what this framework predicts that the rival does not. That is the largest remaining weakness in this document, and it is not repaired by adding citations at the moment of registration, because a reference list assembled to look complete is worth less than a short one that says what it is missing. The commitment registered instead: before any prediction above is scored, the external work each prediction stands against is cited in the transparent-changes record or in a registered successor, and any prediction whose surrounding literature has not been engaged by that point is scored untested rather than supported, whatever its own measurement reports.

ENGAGEMENT OF THE EXTERNAL WORK P21 STANDS AGAINST, ADDED BY SUCCESSOR VERSION 1.103, AS THE COMMITMENT ABOVE REQUIRES BEFORE P21 IS SCORED.

Engaged on 11 October 2026. Each work below was read on that day at the official address given, is described in its own words and is granted no more than it states. Version 1.102 named no external work against P21, so this is the proposition's first engagement; it changes none of P21's words, its refuter or its confidence.

THE CLOSEST PRIOR CLAIM. No published work found states P21 itself: that correction taking part in each round pulls away, as depth rises, from correction that sees only the finished output. Two claims come nearest to its shape. Xie and colleagues (2023) put a step-by-step self-evaluation inside the decoding of a reasoning chain and reported that "performance gains (in absolute accuracy % increase) increase as reasoning chains become longer on both GSM8K and StrategyQA": against the same prompting without it, the gain ran plus 5.49, plus 7.82 and plus 9.78 percentage points from the shortest chains to the longest on the first of these benchmarks and plus 0.31, plus 0.46 and plus 3.42 on the second, and where the method fell behind a baseline of equal cost they read it as "the limitation of our method when applied to shorter reasoning chains". That is a gain over prompting with no step-level evaluation, which is this proposition's control and not its contrast; depth there is the length of chain a problem needs, so it is not separated from the problem's difficulty; and the figures carry no interval. Zhang (2026), a single-author preprint first posted on 23 March 2026, which its arXiv record states was accepted at an ICLR 2026 workshop, is the nearest to the contrast itself. Over five rounds of self-improvement training of one model on one dataset, it compared keeping each round's self-generated solutions with no check, with a check of the final answer, and with that check together with checks of each step's arithmetic and consistency, which it calls VSI, and it reports that "VSI keeps improving over all 5 rounds, while no verification collapses and outcome-only filtering plateaus." Its own table puts the step-level arm's lead over the answer-level arm at minus 0.6, plus 0.5, plus 1.6, plus 3.2 and plus 5.2 percentage points in rounds one to five, and its table for a second test set at plus 0.8, plus 1.4, plus 2.8, plus 4.0 and plus 5.0: a lead that grows with the rounds. It gives no interval and no test of that trend, it notes variation of one to two

percentage points between evaluation runs, and it says itself that "How stable each approach is over many rounds remains to be confirmed." Its two checking arms act alike, by keeping or discarding a solution, so the action space this document names is not varied in it; its recursion is retraining, not the revision of an artefact; and it traces the stall of its answer-level arm to "standard filtering criteria that retain solutions based solely on final answer correctness, which lets lucky guesses (correct answers with flawed reasoning) contaminate the training data".

The advantage without a widening in depth is older and better attested. Lightman and colleagues (2023) found that "process supervision significantly outperforms outcome supervision for training models to solve problems from the challenging MATH dataset", and, comparing a reward model trained on step labels (PRM) with one trained on outcomes, that "Not only does the PRM reach higher performance for all values of N, but the performance gap widens as N increases." N there counts the candidate solutions searched, so that widening is in the breadth of search, and across difficulty "the performance gap is not only apparent on high difficulty problems: it is in fact apparent across all difficulties." Uesato and colleagues (2022) found that "pure outcome-based supervision produces similar final-answer error rates with less label supervision", and that "for correct reasoning steps we find it necessary to use process-based supervision or supervision from learned reward models that emulate process-based feedback". Tyen and colleagues (2023), in a paper published in the Findings of ACL 2024 as its arXiv record states, report that "we show that poor self-correction performance stems from LLMs' inability to find logical mistakes, rather than their ability to correct a known mistake"; testing correction apart from the finding of mistakes, "using a backtracking setup that feeds ground truth mistake location information to the model", they found that "this boosts downstream task performance across our 5 reasoning tasks". That is a level result: models that generally struggle to find mistakes can correct them given their location, with no depth, no rounds and no corrector limited to accepting or rejecting. It is still the nearest prior claim found to the information asymmetry this proposition names, "whether the corrector sees the process or only its output" (line 1256 of version 1.102): what it reports to be hard is finding the mistake, which a corrector that sees only the finished output must do. Where a monitor or corrector should sit has been argued in words. Lindner and colleagues (2025), in an analysis with case studies and no experiment, set out three forms of review, from one that blocks all action while anything flagged awaits review to one that never blocks, and hold that "Fully-synchronous monitoring provides the strongest safety properties, but has prohibitive latency for large-scale deployments", the form that never blocks failing where harm can be done before review, or where partial progress is irreversible and a single failed attempt does not end the deployment. Jia and colleagues (2026), writing of financial systems that call tools, say that filters at the boundary of a conversation and judges applied after a run both stand outside it, and that "an effective defense must instead operate inline within the execution loop itself"; their evaluation, which they call single-run and benchmark-bound, sets their method against two prompt baselines, per-step judges, a firewall and two adapted guardrail methods, and not against a judge applied after the run. Neither paper states a widening with depth.

A gap that grows with the horizon, between correction on a learner's own path and supervision given outside it, is a formal result in imitation learning. Ross and Bagnell (2010) showed that training on demonstrations alone "leads to compounding errors and a regret bound that grows quadratically in the time horizon of the task", and that two alternatives, in which training runs over

episodes of interaction as the learner's own policy is phased in, give "stronger performance guarantees"; Ross, Gordon and Bagnell (2011) extended the remedy to any no-regret learner. The theory of noisy computation holds a result of the same shape: Aharonov and colleagues (1996) proved that a noisy reversible circuit whose inputs are all present at the start "must have size exponential in its depth in order to compute a function with high probability", and drew from it that "fault tolerant QC must use a constant supply of inputs all the time", QC being a quantum circuit. Those results concern the training of a policy and noise in circuits, not a corrector revising an artefact or a monitor reading a finished output, and this entry claims no priority over that shape in those settings. The mechanism registered beside P21's interaction has been stated before as an expectation and not as a measurement: Kutasov and colleagues (2025) wrote that "our measurements already demonstrate the difficulty of monitoring for subtle sabotage attempts, which we expect to only increase in the face of more complex and longer-horizon tasks."

THE CLOSEST EXISTING APPARATUS. No published design found crosses where correction is placed with recursive depth and reads the interaction as its endpoint; the nearest designs each hold part of it. Three bear on placement. Kutasov and colleagues (2025) set a monitor shown a long transcript after every turn against one shown it once at the end, and report that "Although turn-by-turn monitoring does slightly improve performance, the monitor ends up flagging many more transcripts as highly suspicious", and of the cost that "We observe that a higher budget for monitoring is better spent on requerying the monitor at the end of the transcript and averaging the results." In an appendix they explain why turn-by-turn monitoring helped one monitor and not another: "turn-by-turn monitoring increases both the false positive rate and the true positive rate by taking the maximum over many suspicion scores". A verdict taken as the highest of many scores thus raises false alarms as well as detections, which bears on how a monitor's detection is read as the artefact it inspects grows. Liu and colleagues (2026) set a monitor and adviser acting during the run of a programming system against a method which, in their words, "regenerates a plan from a completed trajectory and reruns the executor" (Hayashi and colleagues, 2025, whose own paper reports gains for it). On one benchmark the method acting during the run resolved 15.30, 10.23 and 7.57 percentage points more of the tasks than the method acting after it, on three executors, and the method acting after it fell below no correction on two of them; the authors put that down to specific but mistaken advice which a weaker executor follows, and their reading by difficulty, in which "these gains are concentrated on harder tasks", is against no correction. Stickland and colleagues (2025) study review after the fact with a model of deployment built on their monitor's measured scores, and find that where one attack suffices and harm is immediate "we get 0% safety. This is because attacks go live before they are audited", most of the safety returning where harm is delayed or several attacks are needed. None of the three reads its contrast against depth, and the survey of Kamoi and colleagues (2024), which names correction after a response against correction at generation time as an axis of the field, says of it only that "Post-hoc correction is more flexible and applicable to broader tasks, although generation-time correction is popular for reasoning tasks".

On access at a single depth, Arnav and colleagues (2025) "compare CoT monitoring to action-only monitoring, where only final outputs are reviewed", CoT being the monitored model's chain of thought, and found that "while CoT monitoring is more effective than overseeing only model outputs in scenarios where action-only monitoring fails to reliably identify sabotage, reasoning

traces can contain misleading rationalizations that deceive the CoT monitors, reducing performance in obvious sabotage cases"; they add that "Our evaluation focused on single-step coding problems in a controlled environment". Baker and colleagues (2025) found that "CoT monitoring can be far more effective than monitoring agent actions and outputs alone", and that with too much optimisation against such a monitor "agents learn obfuscated reward hacking, hiding their intent within the CoT while still exhibiting a significant rate of reward hacking", which bears on the trade-off between scaling and independence that line 1260 of version 1.102 leaves open; Guan and colleagues (2025) found again that "CoT monitoring is more effective than action-only monitoring in practical settings". For checkers of reasoning, Miao and colleagues (2023) set a checker of the whole solution against checking step by step and conclude that "This lack of ability to deal with the difficulty of global checking is what makes step checking necessary", and Ling and colleagues (2023) that "directly verifying the validity of an entire deductive reasoning process is challenging". On the size of the monitor's view, Kale and colleagues (2025) set monitors that read a long trajectory in chunks, in order or in parallel, against the baseline of prompting a model with the whole trajectory at once, in a task where "Finding the harmful actions amounts to a needle-in-the-haystack reasoning problem", and report that "the hybrid scaffolding consistently outperforms baseline monitor scaffolding". On the action space, Greenblatt and colleagues (2023) set a protocol in which a weaker trusted model replaces suspicious code with its own solution against one in which it edits that code, and report that "Trusted editing improves beyond just using trusted monitoring and defer to trusted." Their table gives trusted editing a safety of 92 per cent at a usefulness of 94 per cent, against 83 per cent at 93 per cent for trusted monitoring with deferral, on problems solved one at a time, where no artefact carries from one problem to the next, though the protocol's audit budget, its free early audits and any shutdown run across the sequence. Bhatt and colleagues (2025) carried such protocols to tasks of up to fifteen steps and wrote that "Important open questions remain about protocol performance with longer trajectories, more diverse tasks, and imperfect auditing."

On correction across rounds, Madaan and colleagues (2023) report gains from iterative self-feedback over one-step generation, noting that "the marginal improvement naturally decreases with more iterations"; Huang and colleagues (2023) found that "LLMs struggle to self-correct their responses without external feedback, and at times, their performance even degrades after self-correction", and Kamoi and colleagues that "self-correction works well in tasks that can use reliable external feedback". Olausson and colleagues (2023), who set the repair of a program against fresh draws at equal cost, found that "when the cost of carrying out repair is taken into account, performance gains are often modest, vary a lot between subsets of the data, and are sometimes not present at all". Stechly and colleagues (2024), who set a model's critique of its own answers against a sound outside verifier in the same loop, observed "significant performance collapse with self-critique and significant performance gains with sound external verification"; they found that "as the number of backprompts increases, this kind of self-correction consistently degrades output quality", a backprompt being a prompt sent back with the criticism, and that "merely re-prompting with a sound verifier maintains most of the benefits of more involved setups". Correction inside long runs raises resolution rates against a system left uncorrected (Gandhi and colleagues, 2025; Liu and colleagues, 2026); in the first of these the gains of the strongest setting are 11.9, 10.7 and 4.4 percentage points on easy, medium and hard tasks, so its larger gains on the harder tasks are relative ones. In self-consuming training loops correction holds off collapse, against no correction (Gillman

and colleagues, 2024; Feng and colleagues, 2024). Zhang's three arms are the only apparatus found that runs a step-level check, an answer-level check and no check side by side across rounds.

THE STRONGEST RIVAL EXPLANATION is the one line 1258 of version 1.102 already names: "At any material rejection rate a widening advantage accumulates with rounds from that difference alone, whether or not the monitor's view degrades, so a supported verdict can be produced by the action space rather than by the information asymmetry this proposition names." The external record supports it from four sides. First, what a corrector may do with what it flags changes the outcome when nothing else changes: at a single step in the editing protocol above; in Stickland and colleagues' model, where one monitor is safe or unsafe according to whether an action can still be stopped; and in Lindner and colleagues' analysis, where review after the fact fails if partial progress cannot be reversed and one failed attempt does not end the deployment. Second, uncorrected error compounds: Ross and Bagnell's quadratic bound arises because a single mistake can take the learner into states its training never covered; Ord (2025) explains success on longer tasks, within one suite of them, by "a constant rate of failing during each minute a human would take to do the task", which implies "an exponentially declining success rate with the length of the task"; and Zhang's own account of its answer-level arm is that a lucky guess enters the training set and "the model picks up the bad arithmetic, which compounds over rounds". Third, a corrector limited to refusing loses little where a refusal is followed by a fresh attempt, as Stechly and colleagues' sound verifier shows; the outside arm of P21's deciding unit carries the previous artefact forward on a refusal and asks for no fresh attempt, which is the limitation line 1258 registers. Fourth, for supervision, Jia and colleagues (2025) prove that "under standard data coverage assumptions, reinforcement learning through outcome supervision is no more statistically difficult than through process supervision, up to polynomial factors in horizon", and suggest that any observed gap "between outcome and process supervision likely stems from algorithmic limitations rather than inherent statistical difficulties". Their bound leaves room for a gap that grows polynomially with the horizon, so it predicts neither a constant lead nor an exponential one; this entry reads it as the rival's general form, in which what separates the arms is what each can do with what it receives, not what it is shown.

A second account is the general form of P21's own mechanism and not a rival to it, and is stated so that a supported result is not over-read: a reader of any kind loses accuracy as its input grows, with or without a loop. The nearest measurement is on finished records of runs: Deshpande and colleagues (2025), who set models to find and place the errors in such records, report that "We find all performance metrics to be anti-correlated with input length", with Pearson correlations of minus 0.379, minus 0.291 and minus 0.296 for their three scores. For retrieval, Hsieh and colleagues (2024) report that "almost all models exhibit large performance drops as the context length increases", Modarressi and colleagues (2025) that "performance degrades significantly as context length increases", and Liu and colleagues (2024) that "performance can degrade significantly when changing the position of relevant information". The relation is not uniform: Guan and colleagues find that "longer CoTs are generally more monitorable", where the added length is the monitored model's own reasoning, and Kale and colleagues that "sequential monitoring benefits from a larger chunk size; however, hierarchical monitoring has an optimal chunk size of 4 to 5". None of these measures the detection of a planted fault in a growing artefact, so they make the registered fall plausible without assuring it; and they mean that a size ladder with no loop would show the

cleaned-view part of any widening, which leaves the part owed to earlier misses, the rival's part, as what the loop alone adds. The honest expectation, kept apart from the hypothesis under test and held with low confidence because no engine's rates are known before the registered run, is that the rival will explain part of any widening found, possibly most of it.

WHAT P21 PREDICTS THAT THE RIVAL DOES NOT. Against every account of a level advantage, which is most of the work above, P21 predicts an interaction, a lead that grows with depth, and a lead that is the same at every depth is scored against it (line 1254 of version 1.102). Against the gains over no correction that already grow with length or difficulty (Xie and colleagues; Liu and colleagues, 2026), it predicts that the lead over an outside monitor grows too, on a fixed task whose depth is administered and not set by the problem. Against the compounding and action-space rival, P21's mechanism predicts two things that rival neither needs nor predicts: that the monitor's detection falls as the artefact it inspects grows, and that the widening survives when the monitor's earlier misses are taken out of its view, which the deciding unit's matched control call measures on the outside arm's cleaned view. The rival predicts a widening carried by those misses and by refusals, with detection on a cleaned view flat; where a widening arrives without the registered fall, line 1258 already requires it to be reported as unexplained by this framework. Against the account in which correction inside the round does not help (Huang and colleagues; Kamoi and colleagues), P21 predicts a lead, and under the deciding unit's rules a corrector that does not act is a regime never reached, not a refutation. Against the finding that a model's own critique degrades its output as rounds accumulate (Stechly and colleagues), P21 predicts the opposite sign, and under this successor's directional reading a narrowing is scored against it.

WHAT IS ALREADY KNOWN, AND WHAT A REGISTERED COUNTED TEST ADDS. Already known: a level advantage for correction that sees the steps or the reasoning over correction that sees only the answer or the output, in mathematical reasoning and in monitoring, with misleading reasoning as the stated exception; in one study, that poor self-correction stems from failing to find mistakes rather than from failing to correct a known one, and that correction given the location of a mistake raises performance; gains for evaluation inside a reasoning chain that grow with the chain's length, and, in one study, extra solutions from correction inside long runs that fall largely on the harder tasks, both against no correction; a level advantage for a corrector acting during a run over one acting after it, in one study, and only a slight one for a monitor consulted at every turn over one consulted at the end, in another; a gap that grows with the horizon where nothing corrects a learner on its own path, in imitation learning, with a like result for noisy circuits; one preprint in which a step-level filter pulls away from an answer-level filter over five rounds, without intervals; the expectation that monitoring grows harder as tasks grow longer, a negative correlation between the length of a finished record and how well it is read, and measured falls in retrieval as contexts grow. Known against P21: a model's own critique can degrade its output as rounds accumulate, and a sound verifier that only accepts or rejects can keep most of the benefit where a refusal is followed by a fresh attempt.

Not yet known, and what a registered counted test of the widening adds: the interaction of placement with depth itself as the registered endpoint, on a fixed task with administered depth, with a tolerance, a precision gate and a refuter that scores a constant lead against P21, and under this

successor's directional reading also no widening or a narrowing, so that the claim can lose, which no work found risks; the split of any widening into the part owed to the monitor's falling detection on a cleaned view and the part owed to faults it cannot undo, which separates P21's mechanism from its strongest rival; and scoring counted by frozen code against ground truth fixed before any call, with no judge. Its reach stays where line 1248 of version 1.102 and the deciding unit put it: an access interaction in an administered revision loop of checkable faults on fixed-weight engines, showing nothing about judged correction, safety, build order, the Eden Protocol as an architecture, or recursion as against a longer artefact.

REFERENCES FOR P21, each read on 11 October 2026 at the address given.

Aharonov, D., Ben-Or, M., Impagliazzo, R. and Nisan, N. (1996). Limitations of Noisy Reversible Computation. arXiv:quant-ph/9611028, 17 November 1996. https://doi.org/10.48550/arXiv.quant-ph/9611028

Arnav, B. et al. (2025). CoT Red-Handed: Stress Testing Chain-of-Thought Monitoring. arXiv:2505.23575, first version 29 May 2025, version 3 of 24 November 2025 read. https://doi.org/10.48550/arXiv.2505.23575

Baker, B. et al. (2025). Monitoring Reasoning Models for Misbehavior and the Risks of Promoting Obfuscation. arXiv:2503.11926, 14 March 2025. https://doi.org/10.48550/arXiv.2503.11926

Bhatt, A. et al. (2025). Ctrl-Z: Controlling AI Agents via Resampling. arXiv:2504.10374, 14 April 2025. https://doi.org/10.48550/arXiv.2504.10374

Deshpande, D. et al. (2025). TRAIL: Trace Reasoning and Agentic Issue Localization. arXiv:2505.08638, first version 13 May 2025, version 3 of 23 June 2025 read. https://doi.org/10.48550/arXiv.2505.08638

Feng, Y. et al. (2024). Beyond Model Collapse: Scaling Up with Synthesized Data Requires Verification. arXiv:2406.07515, first version 11 June 2024, version 2 of 25 October 2024 read. https://doi.org/10.48550/arXiv.2406.07515

Gandhi, S. et al. (2025). When Agents go Astray: Course-Correcting SWE Agents with PRMs. arXiv:2509.02360, first version 2 September 2025, version 2 of 21 October 2025 read. https://doi.org/10.48550/arXiv.2509.02360

Gillman, N. et al. (2024). Self-Correcting Self-Consuming Loops for Generative Model Training. arXiv:2402.07087, first version 11 February 2024, version 3 of 10 June 2024 read. https://doi.org/10.48550/arXiv.2402.07087

Greenblatt, R., Shlegeris, B., Sachan, K. and Roger, F. (2023). AI Control: Improving Safety Despite Intentional Subversion. arXiv:2312.06942, first version 12 December 2023, version 5 of 23 July 2024 read. https://doi.org/10.48550/arXiv.2312.06942

Guan, M. Y. et al. (2025). Monitoring Monitorability. arXiv:2512.18311, 20 December 2025. https://doi.org/10.48550/arXiv.2512.18311

Hayashi, H. et al. (2025). Self-Abstraction from Grounded Experience for Plan-Guided Policy Refinement. arXiv:2511.05931, 8 November 2025. https://doi.org/10.48550/arXiv.2511.05931

Hsieh, C.-P. et al. (2024). RULER: What's the Real Context Size of Your Long-Context Language Models? arXiv:2404.06654, first version 9 April 2024, version 3 of 6 August 2024 read. https://doi.org/10.48550/arXiv.2404.06654

Huang, J. et al. (2023). Large Language Models Cannot Self-Correct Reasoning Yet. arXiv:2310.01798, first version 3 October 2023, version 2 of 14 March 2024 read. https://doi.org/10.48550/arXiv.2310.01798

Jia, Z., Rakhlin, A. and Xie, T. (2025). Do We Need to Verify Step by Step? Rethinking Process Supervision from a Theoretical Perspective. arXiv:2502.10581, first version 14 February 2025, version 2 of 26 March 2025 read. https://doi.org/10.48550/arXiv.2502.10581

Jia, H. et al. (2026). FinHarness: An Inline Lifecycle Safety Harness for Finance LLM Agents. arXiv:2605.27333, 26 May 2026. https://doi.org/10.48550/arXiv.2605.27333

Kale, N. et al. (2025). Reliable Weak-to-Strong Monitoring of LLM Agents. arXiv:2508.19461, 26 August 2025. https://doi.org/10.48550/arXiv.2508.19461

Kamoi, R. et al. (2024). When Can LLMs Actually Correct Their Own Mistakes? A Critical Survey of Self-Correction of LLMs. Transactions of the Association for Computational Linguistics 12, 1417 to 1440; arXiv:2406.01297, first version 3 June 2024, version 3 of 3 December 2024 read. https://doi.org/10.1162/tacl_a_00713 and https://aclanthology.org/2024.tacl-1.78/ and https://doi.org/10.48550/arXiv.2406.01297

Kutasov, J. et al. (2025). SHADE-Arena: Evaluating Sabotage and Monitoring in LLM Agents. arXiv:2506.15740, first version 17 June 2025, version 2 of 8 July 2025 read. https://doi.org/10.48550/arXiv.2506.15740

Lightman, H. et al. (2023). Let's Verify Step by Step. arXiv:2305.20050, 31 May 2023. https://doi.org/10.48550/arXiv.2305.20050

Lindner, D. et al. (2025). Practical challenges of control monitoring in frontier AI deployments. arXiv:2512.22154, 15 December 2025. https://doi.org/10.48550/arXiv.2512.22154

Ling, Z. et al. (2023). Deductive Verification of Chain-of-Thought Reasoning. arXiv:2306.03872, first version 6 June 2023, version 3 of 3 October 2023 read. https://doi.org/10.48550/arXiv.2306.03872

Liu, N. F. et al. (2024). Lost in the Middle: How Language Models Use Long Contexts. Transactions of the Association for Computational Linguistics 12, 157 to 173; arXiv:2307.03172, first version 6 July 2023, version 3 of 20 November 2023 read. https://doi.org/10.1162/tacl_a_00638 and https://aclanthology.org/2024.tacl-1.9/ and https://doi.org/10.48550/arXiv.2307.03172

Liu, S. et al. (2026). Online Monitoring and Corrective Steering of Programming Agents. arXiv:2608.06701, 7 August 2026. https://doi.org/10.48550/arXiv.2608.06701

Madaan, A. et al. (2023). Self-Refine: Iterative Refinement with Self-Feedback. arXiv:2303.17651, first version 30 March 2023, version 2 of 25 May 2023 read. https://doi.org/10.48550/arXiv.2303.17651

Miao, N., Teh, Y. W. and Rainforth, T. (2023). SelfCheck: Using LLMs to Zero-Shot Check Their Own Step-by-Step Reasoning. arXiv:2308.00436, first version 1 August 2023, version 3 of 5 October 2023 read. https://doi.org/10.48550/arXiv.2308.00436

Modarressi, A. et al. (2025). NoLiMa: Long-Context Evaluation Beyond Literal Matching. arXiv:2502.05167, first version 7 February 2025, version 3 of 9 July 2025 read. https://doi.org/10.48550/arXiv.2502.05167

Olausson, T. X. et al. (2023). Is Self-Repair a Silver Bullet for Code Generation? arXiv:2306.09896, first version 16 June 2023, version 5 of 2 February 2024 read. https://doi.org/10.48550/arXiv.2306.09896

Ord, T. (2025). Is there a half-life for the success rates of AI agents? arXiv:2505.05115, 8 May 2025. https://doi.org/10.48550/arXiv.2505.05115

Ross, S. and Bagnell, J. A. (2010). Efficient Reductions for Imitation Learning. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR 9, 661 to 668. https://proceedings.mlr.press/v9/ross10a.html

Ross, S., Gordon, G. J. and Bagnell, J. A. (2011). A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, PMLR 15, 627 to 635; arXiv:1011.0686, first version 2 November 2010, version 3 of 16 March 2011 read. https://proceedings.mlr.press/v15/ross11a.html and https://doi.org/10.48550/arXiv.1011.0686

Stechly, K., Valmeekam, K. and Kambhampati, S. (2024). On the Self-Verification Limitations of Large Language Models on Reasoning and Planning Tasks. arXiv:2402.08115, first version 12 February 2024, version 2 of 3 August 2024 read. https://doi.org/10.48550/arXiv.2402.08115

Stickland, A. C. et al. (2025). Async Control: Stress-testing Asynchronous Control Measures for LLM Agents. arXiv:2512.13526, 15 December 2025. https://doi.org/10.48550/arXiv.2512.13526

Tyen, G. et al. (2023). LLMs cannot find reasoning errors, but can correct them given the error location. arXiv:2311.08516, first version 14 November 2023, version 3 of 4 June 2024 read. https://doi.org/10.48550/arXiv.2311.08516

Uesato, J. et al. (2022). Solving math word problems with process- and outcome-based feedback. arXiv:2211.14275, 25 November 2022. https://doi.org/10.48550/arXiv.2211.14275

Xie, Y. et al. (2023). Self-Evaluation Guided Beam Search for Reasoning. arXiv:2305.00633, first version 1 May 2023, version 3 of 26 October 2023 read. https://doi.org/10.48550/arXiv.2305.00633

Zhang, X. (2026). Reliable Self-Improvement Training by Verifying Reasoning, Not Just Answers. arXiv:2603.21558, first version 23 March 2026, version 2 of 28 May 2026 read.

https://doi.org/10.48550/arXiv.2603.21558

This entry discharges the commitment above for P21 alone.

ENGAGEMENT OF THE EXTERNAL WORK P17 STANDS AGAINST, ADDED BY SUCCESSOR VERSION 1.103, AS THE COMMITMENT ABOVE REQUIRES BEFORE P17 IS SCORED.

Engaged on 11 October 2026. Each work below was read on that day at the official address given, except the three papers in IEEE Transactions on Software Engineering, which were read at the publisher's record in the IEEE Computer Society Digital Library because the page their address resolves to answered only with an automated challenge; each is described in its own words and is granted no more than it states. Version 1.102 named two works against P17 (lines 1484 and 1490) and three studies of multiversion software against P6, the bound that P17's premise implies (lines 1472 to 1474), and engaged none of them; this is the proposition's first engagement, and it changes none of P17's words, its refuter or its confidence.

THE CLOSEST PRIOR CLAIM. No published work found claims that correctors drawn from capable models miss faults independently of one another; every measurement found reports the opposite. The claim nearest P17's own is the assumption that the literature on multiversion software named and then tested. Knight and Leveson (1986) state that N-version programming "depends for its reliability improvement on the assumption that programs that have been developed independently will fail independently"; they had 27 versions of one program prepared independently from the same specification at two universities, subjected them to one million tests, and found that "the number of tests in which more than one program failed was substantially more than expected". Eckhardt and Lee (1985) give a theoretical basis for analysing such software, in which an intensity function "describes the propensity of programmers to introduce design faults in such a way that software components fail together when executing in the application environment". Littlewood and Miller (1989) read that work as showing "that independently developed program versions fail dependently", and show that diverse methodologies of development lower the probability that several versions fail together, so that "it is theoretically possible to obtain versions which exhibit better than independent failure behavior". The same assumption is named for the judges whose verdicts are pooled in the evaluation of language models: most of the classical methods of pooling "assume annotators are conditionally independent given the true label", in the words of Balasubramanian and colleagues (2026), "an assumption often violated by LLM judges due to shared data, architectures, prompts, and failure modes".

Among language models, every measurement found points against independence. Kim and colleagues (2025) evaluate over 350 models on two leaderboards and a resume-screening task and report that "on one leaderboard dataset, models agree 60% of the time when both models err", and that "larger and more accurate models have highly correlated errors, even with distinct architectures and providers"; their measure asks "do models agree on an answer when both err", which is agreement on the wrong answer and not the rate at which two models are wrong together. Goel and colleagues (2025) find that "model mistakes are becoming more similar with increasing capabilities, pointing to risks from correlated failures". Kohli (2026) finds that nine judges from seven model families "effectively provide only about 2 independent votes' worth of information", because "the

models make the same mistakes on the same items"; Hossain and colleagues (2026) find an average pairwise correlation between judge errors of 0.21 in a bank of ten judges, worth "roughly as much statistical information as 3.5 independent judges", with the dependency "even stronger among the high-accuracy frontier judges we evaluate, including judges from different providers". Spiro (2026) finds the forecasting errors of three systems from different developers correlated at a mean of 0.77, and concludes that "three nominally independent AI systems share the same failure modes". Bugaud (2026) finds, across vision-language models, that "models from the same architectural family share correlated errors that standard voting ignores". In generated code, Nogueira and colleagues (2026) find implementations from different models "still failing on the same tests far more often than expected under independence", and Ron and colleagues (2026), evaluating 48 implementations written by coding systems from the specification Knight and Leveson used, on a shared oracle and 1,000,000 randomised test inputs, report "substantial common-mode failure". Bhardwaj and colleagues (2026) find, in a preregistered evaluation scored by deterministic code, that two instances of one model "co-fail on 90.0% of the missions on which either fails". The nearest claim in the other direction is a practice and not a measurement of independence: Verga and colleagues (2024) report that a panel of smaller models "exhibits less intra-model bias due to its composition of disjoint model families".

THE CLOSEST EXISTING APPARATUS. Two kinds of design come near the deciding unit, and none found reads independence as an equivalence claim inside a margin fixed before the data. The first kind measures how far two decision makers err on the same items beyond what chance would give at their accuracies, which is the baseline P17's refuter reads against, the rate of joint misses that independence predicts from the two miss rates (line 1148). Geirhos and colleagues (2020) introduce "trial-by-trial error consistency, a quantitative analysis for measuring whether two decision making systems systematically make errors on the same inputs", and find that "Irrespective of architecture, CNNs are remarkably consistent with one another". Goel and colleagues build on it a chance-adjusted measure, "a metric for LM similarity based on overlap in model mistakes", from error consistency, which in their words "measures overlap in the samples where two models err beyond what is expected by chance". Kohli reads a panel against "a Condorcet null model" with the Kish effective sample size, and Hossain and colleagues read one by the average pairwise correlation of judge errors and the number of independent judges it is worth. Chen (2026) states a limit of "the usual diagnostic, average pairwise error correlation rho", which "cannot identify beta", that is "the rate at which every model is wrong on the same query", because "error laws with identical marginals and pairwise correlations can have different all-wrong rates"; by that reason the limit holds for any measure read from the pairs alone. The second kind tests the assumption of independence directly. The experiment of Knight and Leveson is its classic form; Nogueira and colleagues propose "the first systematic methodology to assess failure independence in LLM-generated code"; and Bhardwaj and colleagues, the nearest design found, test the assumption behind compositional reliability bounds, which multiply the reliabilities of components in what they call "a step licensed by a conditional-independence assumption that is routinely stated and rarely tested", in a preregistered evaluation with deterministic scoring that manipulates whether two components share a model, and report that "Every interval excludes independence by a wide margin". Their design differs from the deciding unit's in what it reads: components carrying out missions under contracts, the confirmatory arrangement a handoff from one component to the next, each component scored for failure and the

pair tested for dependence; the deciding unit reads correctors reviewing the same planted faults separately, each scored for a catch or a miss, with independence the claim that must show itself inside the margin.

THE STRONGEST RIVAL EXPLANATION is the one line 1156 of version 1.102 already names: shared difficulty. The deciding unit's "own registered operating characteristics show the marginal reading firing under heterogeneity in fault difficulty shared by both panels with no substrate effect present" (line 1156). The arithmetic is short. Where two correctors miss each fault independently of one another given the fault, their joint-miss rate over the set equals the product of their miss rates plus the covariance of their per-fault miss probabilities. For two correctors whose miss probabilities rise and fall in proportion across the faults, and so share one coefficient of variation, the ratio of the joint-miss rate to that product is one plus the square of that coefficient, so the ratio passes the registered margin of 0.20 on the log scale once the coefficient exceeds about 0.47. No shared substrate is needed for this, only faults that are harder for every corrector. The external record supports the rival. Ron and colleagues trace most of the common-mode failures they found to a few places: "the failures are driven by a small number of difficult or ambiguous parts of the specification". Littlewood and Miller find "a precise duality between input choice and program choice" in the model of Eckhardt and Lee. A second rival is the account the framework's premise is about, shared substrate: correctors that share training data, architecture or provider share blind spots. Bommasani and colleagues (2022) state it as a hypothesis, "if decision-makers share components like training data or specific models, then they will produce more homogeneous outcomes", and find that "sharing training data reliably exacerbates homogenization", with "mixed results" for one shared model adapted across tasks; Kim and colleagues identify "shared architectures and providers" among the factors driving correlation; Ghanem (2026) argues from "A coincident-failure model in the tradition of Eckhardt and Lee" that organisational independence "no longer implies statistical independence once roles share a substrate"; and in the experiment of Bhardwaj and colleagues, substituting a different model "reduces the association in six of six contrasts", while substituting a different vendor, with the model already different, does not, "a registered hypothesis reported as a null". Both rivals predict a refutation of P17, and line 1156 requires the reading that tells them apart to be printed beside any refutation.

WHAT P17 PREDICTS THAT THE RIVAL DOES NOT. Both rivals predict that some pair of correctors in a panel misses together more often than the product of its miss rates, by more than the margin, and P17 predicts that no pair does. They part from P17, and from each other, where the deciding unit can see it. Any excess of a pair's joint misses over that product is bounded by how far each corrector's own misses vary from fault to fault, which shows in how often a corrector misses a fault again when it reviews the same fault a second time; P17 is read only where every pair could have shown a correlation of the registered size (line 1150), so a supported result cannot come from faults that are equally hard for every corrector or from misses that strike at random. Where each corrector's own misses do recur, P17 predicts that the faults one corrector finds hard are not the faults the other finds hard, so that every pair stays inside the margin; shared difficulty predicts that they are the same faults, raising pairs within one engine and pairs across engines alike; shared substrate predicts pairs within one engine above the margin and above the pairs across engines, which is the contrast the attribution reading measures (line 1156). The honest expectation, from

every measurement above, is that a rival will fit: a refutation is what the external record leads one to expect, and a supported result is the outcome that would depart from it.

WHAT IS ALREADY KNOWN, AND WHAT A REGISTERED COUNTED TEST ADDS. Already known: that independently developed programs fail together more often than independence predicts, by experiment and in theory, and that diverse methods of development can in theory do better than independence; that the errors of language models are correlated across models, architectures and providers, more strongly as accuracy rises, in leaderboard questions, forecasts, the verdicts of judges and generated code; that panels of model judges are worth far fewer independent votes than their size; that two components sharing a model co-fail far more often than independence predicts, in a preregistered comparison scored by deterministic code; that common failures concentrate where a specification is hard or ambiguous; and that pairwise measures cannot fix the rate at which every member of a panel fails at once. Known against P17: every measurement found. Known for it: only the theoretical possibility, under forced diversity of method, of better than independent behaviour.

Not yet known, and what a registered counted test of P17 adds: independence read as an equivalence claim inside a margin fixed before the data, so that it is supported only by showing itself inside that margin and a design too weak to show it returns inconclusive; panels named in advance by attributes fixed before any fault is scored, among them the population most likely to share blind spots, correctors drawn from one engine; a fault population authored and calibrated on a corrector outside both panels, its labels fixed by frozen code before any review; a demonstration that every pair could have shown the registered correlation; a refutation banked only when it recurs on a second sealed fault set; and the attribution reading beside any refutation, which separates shared difficulty from shared substrate; all counted by frozen code, with no judge. Its reach stays where lines 1142, 1144 and 1346 of version 1.102 and the deciding unit put it: coincident misses on an authored population of checkable faults, by correctors that are fixed-weight engines under registered procedures; a cross-sectional dependence that measures no temporal accumulation and no exponent; and at most one component of the condition set from which one half is derived, never the value two.

REFERENCES FOR P17, each read on 11 October 2026 at the address given, except the three papers in IEEE Transactions on Software Engineering, read that day at the publisher's record in the IEEE Computer Society Digital Library.

Balasubramanian, K., Podkopaev, A. and Kasiviswanathan, S. P. (2026). Dependence-Aware Label Aggregation for LLM-as-a-Judge via Ising Models. arXiv:2601.22336, 29 January 2026. https://doi.org/10.48550/arXiv.2601.22336

Bhardwaj, V. P., Singh, G. and Bhardwaj, A. P. (2026). Agent Behavioral Contracts II: Certifying Compositional Reliability Without Assuming Independence. arXiv:2608.12895, 13 August 2026. https://doi.org/10.48550/arXiv.2608.12895

Bommasani, R. et al. (2022). Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization? arXiv:2211.13972, 25 November 2022, published at NeurIPS 2022 as its arXiv record states. https://doi.org/10.48550/arXiv.2211.13972

Bugaud, Z. (2026). Hidden Clones: Exposing and Fixing Family Bias in Vision-Language Model Ensembles. arXiv:2603.17111, 17 March 2026. https://doi.org/10.48550/arXiv.2603.17111

Chen, J. (2026). When Does Combining Language Models Help? A Co-Failure Ceiling on Routing, Voting, and Mixture-of-Agents Across 67 Frontier Models. arXiv:2606.27288, 25 June 2026. https://doi.org/10.48550/arXiv.2606.27288

Eckhardt, D. E. and Lee, L. D. (1985). A Theoretical Basis for the Analysis of Multiversion Software Subject to Coincident Errors. IEEE Transactions on Software Engineering SE-11(12), 1511 to 1517. https://doi.org/10.1109/TSE.1985.231895

Geirhos, R., Meding, K. and Wichmann, F. A. (2020). Beyond accuracy: quantifying trial-by-trial behaviour of CNNs and humans by measuring error consistency. arXiv:2006.16736, first version 30 June 2020, version 3 of 18 December 2020 read, the camera-ready version for NeurIPS 2020 as its arXiv record states. https://doi.org/10.48550/arXiv.2006.16736

Ghanem, M. C. (2026). Builder, Defender, Breaker: Measurable Independence and Bounded Autonomy When Generative Models Build, Defend and Test Software. arXiv:2607.03215, first version 3 July 2026, version 2 of 8 September 2026 read. https://doi.org/10.48550/arXiv.2607.03215

Goel, S. et al. (2025). Great Models Think Alike and this Undermines AI Oversight. arXiv:2502.04313, first version 6 February 2025, version 2 of 12 June 2025 read. https://doi.org/10.48550/arXiv.2502.04313

Hossain, E., Yousefi, N. and Lim, S.-N. (2026). Agreement Overstates Evidence: Error Dependence in LLM Judge Consensus. arXiv:2609.22512, 18 September 2026. https://doi.org/10.48550/arXiv.2609.22512

Kim, E. et al. (2025). Correlated Errors in Large Language Models. arXiv:2506.07962, 9 June 2025, accepted to ICML 2025 as its arXiv record states. https://doi.org/10.48550/arXiv.2506.07962

Knight, J. C. and Leveson, N. G. (1986). An experimental evaluation of the assumption of independence in multiversion programming. IEEE Transactions on Software Engineering SE-12(1), 96 to 109. https://doi.org/10.1109/TSE.1986.6312924

Kohli, G. (2026). Nine Judges, Two Effective Votes: Correlated Errors Undermine LLM Evaluation Panels. arXiv:2605.29800, 28 May 2026. https://doi.org/10.48550/arXiv.2605.29800

Littlewood, B. and Miller, D. R. (1989). Conceptual modeling of coincident failures in multiversion software. IEEE Transactions on Software Engineering 15(12), 1596 to 1614. https://doi.org/10.1109/32.58771

Nogueira, R. P. et al. (2026). A Systematic Methodology for Evaluating Failure Independence in LLM-Generated Code. arXiv:2607.02808, 2 July 2026. https://doi.org/10.48550/arXiv.2607.02808

Ron, J., Baudry, B. and Monperrus, M. (2026). N-Version Programming with Coding Agents. arXiv:2606.20158, 18 June 2026. https://doi.org/10.48550/arXiv.2606.20158

Spiro, T. (2026). The Oracle's Fingerprint: Correlated AI Forecasting Errors and the Limits of Bias Transmission. arXiv:2605.00844, 7 April 2026. https://doi.org/10.48550/arXiv.2605.00844

Verga, P. et al. (2024). Replacing Judges with Juries: Evaluating LLM Generations with a Panel of Diverse Models. arXiv:2404.18796, first version 29 April 2024, version 2 of 1 May 2024 read. https://doi.org/10.48550/arXiv.2404.18796

This entry discharges the commitment above for P17 alone, as the entry before it does for P21. Every proposition other than those two remains under it, and one whose surrounding literature has not been engaged before it is scored is scored untested rather than supported, as registered above.

FIVE ASSERTIONS ABOVE CARRY NO IDENTIFIER, and the status is marked rather than blurred. The three reversals disclosed in the background, the unrequested run of 11 August 2026 disclosed there, and the finding that unblinded scoring within a single model family can reverse an alignment result, are claims about the record that a reader cannot currently check from this document alone. Each identifier is carried by the transparent-changes record as it is minted, on the same terms. Until that happens those five are the author's testimony rather than citable record. That is weaker standing than anything else in this document, and it is stated here in those words rather than left for a reader to work out.

SUCCESSOR QUALIFICATION. Three of those five now carry identifiers a reader can follow, each a public component of this programme, checked at OSF on 15 October 2026: the retraction of the estimate of approximately 2.24 is recorded in Paper II (DOI 10.17605/OSF.IO/8FJMA), whose title names the retraction and whose text reads "The initial single-model estimate (alpha approximately 2.24, quadratic) is retracted"; the correction of the ceiling relation on 16 August 2026 is printed with its correction note beside Law III in the statement paper (DOI 10.17605/OSF.IO/GW5MX, version 7.3, "Correction and derivation (v4.0, 16 August 2026)"); and the finding on unblinded scoring is Paper IV.d (DOI 10.17605/OSF.IO/2S3E6), read with the qualification beside condition A above. The downgrade is lifted for those three and for them alone, on the terms it is lifted below for the commits of 16 and 17 March 2026. The withdrawal of the self-acceleration expression and the unrequested run of 11 August 2026 carry no identifier here and stay the author's testimony until one is minted.

FURTHER PUBLIC ANCHORS, EACH DATED.

The Priority Record section of Paper III puts a public date of 13 February 2026 on the programme's named predictions and attaches a falsification criterion to each, F4, F7 and F10 to F12 among them. Two are worth naming: the composition-operator forward prediction, and the leaf venation d/(d+1) exponent, whose correction of 10 March 2026 carries a date of its own. The page is https://michaeldariuseastwood.com/research/papers/paper-iii-alignment-scaling-problem and the date sits in that section. The prediction register on the site is https://michaeldariuseastwood.com/priority-claims and its machine-readable counterpart is https://michaeldariuseastwood.com/research/priority.json. Both resolved when checked on 1 September 2026. The programme's public laboratory notebook (commenced 10 March 2026) is named here without an identifier a reader can follow from this document, and until an identifier is added by amendment it should be read as the author's testimony and not as citable record, on the

same standing as the five assertions marked in the References field. The minute-resolution commits of 16 and 17 March 2026 now carry identifiers a reader can follow: the repository is https://github.com/MichaelDariusEastwood/arc-principle-validation and the commits are 23ce10a, 5435e77, 8edaec0, e5c22dd, f5e06a8 and c853277, so the downgrade is lifted for those commits and for them alone.