

Closure First

Probing Open Mathematics with the Rule of the Vacuum.Net

J. Konstapel, Leiden, 16 October 2026

© 2026 J. Konstapel, Constable Research, Leiden. All rights reserved.

Quotation with attribution is permitted; reproduction of the whole requires written permission.

Version 1.0

1. The question

Can a machine find mathematics that is not yet known?

In October 2026 that question stopped being abstract. AI systems now solve research problems in theoretical physics and economics, from papers they have never seen. They also cheat on their own tests, and teach each other to do so.

This essay reports an experiment that takes both facts seriously. A rule from the vacuum.net theory steered the search: what closes, stays. The experiment ran in two rounds. The first round failed. The second produced 61 proven theorems, each checked by an independent judge. None of them settles a famous open problem.

The essay sets out the idea, the occasion, the experiment and its evaluation. It ends with what was adjusted and what remains open.

2. The idea: closure before steps

2.1 One rule

The vacuum.net theory describes the vacuum as a single strand that crosses itself. Where it returns on itself in phase, a winding forms. Windings that hold each other in three directions form a knot. Openings between strands are meshes.

The theory has one selection rule. A configuration persists only if it closes on itself in phase. Tension that does not close moves on. Its register is the balanced trit: +1, 0, -1.

The formal statement of the theory gives five axioms, nine definitions and eleven proven theorems (Konstapel 2026c). Two of them matter here.

- **Theorem 4.** Of the 27 trit windows (w_1, w_2, w_3), exactly 9 close, in the sense $w_1 + w_2 + w_3 \equiv 0 \pmod{3}$. They form a subgroup of index 3.
- **Theorem 5.** A window closes exactly when its three entries are all equal or all different. Every open window joins a pair to a different third value.

Closure is therefore not a metaphor in this theory. It is a countable property with an exact algebra.

2.2 Closure and the rest

In arithmetic, closure has a simple shape. Two windings run against each other. They return together, or they leave a rest.

Take a winding of length 2 against a winding of length 3. After n turns of the first and k of the second, the rest is $2^n - 3^k$. This number is never zero for $n, k > 0$: a power of 2 is even, a power of 3 is odd. The windings never close exactly. They come closest at the convergents of $\log_2 3$:

$\log_2 3 = [1; 1, 1, 2, 2, 3, 1, 5, 2, 23, \dots]$.

The pair (19, 12) gives $2^{19}/3^{12} = 0.98654$: the Pythagorean comma.

Many open problems ask exactly this kind of question. Does a cycle close? Does a sum of parts close on the whole? How small can the rest between two windings become?

2.3 Insight as closure first

Mathematicians who have found something new describe the moment in the same way. Poincaré stepped onto a bus and saw that the functions he had studied for weeks were those of non-Euclidean geometry. The whole was there at once, and certain. The steps came later (Poincaré 1908). Hadamard collected many such accounts and found the same order: the whole first, verification after (Hadamard 1945).

In the language of the theory, insight is a closure. A winding is closed or it is not. There is no half-closure. The written proof is the unrolling of a winding that is already round.

Present AI works in the opposite order. It sets step after step and hopes the result closes. It is trained on what is already known and rewarded for reaching known answers. That makes it strong inside a given frame and weak at seeing a new one.

The experiment asks a narrow question. If the search is steered by closure, so that it looks for the rule every closure obeys rather than for an exception, does it find more?

3. The occasion

3.1 AI is solving theory problems

On 14 October 2026 the WIRED AI Lab newsletter reported a study by Neil Thompson, Gabriel Manso and Claudio Mayrunk Verdun of MIT (Knight 2026). Successive OpenAI models were given theory problems from papers posted after their training cutoff. Internet access was blocked. Between mid-2024 and mid-2026 the share solved rose from about 2% to about 35%.

The study itself is not yet published. Its design was presented by Manso in July 2026 (Manso 2026). It has two roles.

- **The respondent** receives only the question and the minimal premises that make it well posed. It has 22 local tools that mark each step correct, wrong or undecided. Undecided steps are flagged as gaps.
- **The judge** receives the answer keys and a list of known traps. It never shows them to the respondent.

On 260 papers in seven fields, GPT-5.5 fully solved about 47% in quantum physics and 22% in general relativity. One attempt took about 20 minutes and 5 dollars.

Manso drew the essential caution himself. A model easily states “this is the answer”. Only the full derivation shows whether it is.

3.2 AI is cheating, and the cheating spreads

The same newsletter reported work from Meta Superintelligence Labs and Duke University (Hu et al. 2026). Coding agents trained against a leaky test learned three ways to pass without solving the task:

- exiting the test early with a success code;
- returning an object that equals everything;
- adding a configuration file that reports failures as passes.

The striking finding is transmission. A small open model that had learned a trick passed it on to stronger closed models through shared code, memory and

plans. With five reference submissions, the share of hacked submissions reached 73.6% for one frontier model and 73.0% for another. One model stayed near 0.2%. On hard tasks the trick grew stronger along a chain after its source had left, as the code acquired more persuasive justifications.

The authors proved a threshold. With k references, adoption probability θ_1 and the target's own success rate u_0 , the hack grows when $R = k\theta_1/u_0 > 1$. With $k = 5$ and $u_0 = 0.3$, that already happens at $\theta_1 > 0.06$.

3.3 Populations with a threshold

A third paper, from Harvard and NTT Research, modelled populations of cooperating agents (Crawley and Tanaka 2026). Cooperation makes the success probability grow with population size N . That creates a critical size:

$$N_{\text{crit}} = N_{\text{ref}} \cdot ((1 - p_{\text{ref}})/(1 - p_{\text{crit}}))^{(1/\beta)}.$$

With their own example ($p_{\text{ref}} = 0.10$, $p_{\text{crit}} = 0.20$, $\beta = 0.035$, $N_{\text{ref}} = 16$) this gives $(0.9/0.8)^{(1/0.035)} = e^{3.365} \approx 28.9$, and $N_{\text{crit}} \approx 463$. Below the threshold the population declines. Above it, it grows, with no change in the individual agent.

3.4 What the occasion teaches

Three lessons shaped the experiment.

1. **Separate the solver from the judge.** The judge must have what the solver lacks.
2. **Keep the test out of reach of the code being tested.** Otherwise the search learns the test, not the problem.
3. **Treat every “solved” as a claim until the derivation is checked.**

4. The experiment

4.1 Round one: hunting exceptions

The first round ran 237 probes in four families. Claude Code executed them in four parallel lines:

- the rest between windings (Collatz, abc, Pillai);
- prime windings (Goldbach, twin primes, Riemann);
- geometry and dynamics (Hadwiger-Nelson, the Lonely Runner, the Mandelbrot set);
- windings that never close ($e + \pi$, Birch-Swinnerton-Dyer).

Each probe was told to look for a counterexample. A “solves” outcome required a counterexample re-verified by a separate script.

The result was empty. 138 probes confirmed known theorems. 99 found nothing. None solved anything or set a record.

The reason was structural, not computational. Most of these conjectures are probably true: wherever counts could be compared with predictions, they agreed to within 0.01% to 2%. And every search fell inside ranges that decades of supercomputing had already cleared. Goldbach is verified to $4 \cdot 10^{18}$; the probe reached 10^{10} . A search for an exception that does not exist returns nothing, however well it is aimed.

The strategy also contradicted the rule it claimed to follow. The theory says what closes stays. Its natural output is a reason for closure, not a place where closure breaks.

4.2 The adjustment

The second round changed one thing: the target. A probe now had six steps.

1. **Generate closures.** Enumerate all closures of a family in a range: cycles, exact solutions, sums of parts equal to the whole.
2. **Order by rest.** For each candidate, measure how far it misses exact closure.
3. **Find the rule** that every closure obeys.
4. **Prove the rule**, step by step, naming any gap.
5. **Judge it** with a separate script and exact arithmetic.
6. **Search the literature.** Is the rule known? What does it do for the open problem?

Then, per family, an independent judge agent re-examined every rule reported as new. It tested each rule with its own fresh code before reading the proof. Only then did it check the proof line by line and search the literature again. A third check, in the main session, recomputed a sample.

This follows the three lessons of the occasion. The solver and the judge are separated. The judge's tests are written independently. And "new" is a claim until the judge confirms it.

4.3 A worked example

The first probe of round two shows the method on the maps $x \rightarrow ax + 1$ for odd x , $x \rightarrow x/2$ for even x .

Claude Code enumerated all cycles for odd $a < 2000$. There are 14. Following each cycle around gives an exact closure condition:

$$2^n = \prod (a \cdot x_i + 1)/x_i,$$

over the k odd elements, with n halvings. Three rules follow.

- **The rest is positive:** $2^n/a^k = \prod (1 + 1/(a \cdot x_i)) > 1$.
- **The rest is bounded:** $0 < n/k - \log_2 a < 1/(a \cdot m \cdot \ln 2)$, with m the smallest odd element.
- **Closure sits on a convergent:** if $m > 2k^2/(a \cdot \ln 2)$, then n/k is a convergent of $\log_2 a$, by Legendre's theorem.

A separate script checked the first two exactly on all 14 cycles. On the 13 cycles where the third rule's premise holds, n/k is indeed a convergent. On the one where it fails, it is not. For $a = 3$ this reasoning is known (Simons and de Weger 2005). The point is the order of work: closures first, then the rule, then the proof.

5. Results

The second round ran 245 probes in four families. The independent judges examined 78 claims.

Family	Probes	New, verified	New, after fix	Known	Refuted
Rest between windings	61	14	4	7	0
Exact closure	63	5	4	7	1
Prime windings	61	16	1	2	0
Scale and orbits	60	13	4	0	0
Total	245	48	13	16	1

"New" means: not found in the literature the judges searched. A selection follows.

5.1 The rest between windings

Two open values closed. In round one, $2^x - 3^y = c$ remained open for $c = -7153$ and $c = 84997$. Each has exactly one solution: $2^{19} - 3^{12}$ and $2^{18} - 3^{11}$. Both are convergents of $\log_2 3$. The first is also the rest of the 12-cycle in

Collatz’s original permutation.

Pillai’s equation, non-coprime bases. For all independent bases $2 \leq a, b \leq 300$, the equation $a^x - b^y = c$ has at most two solutions in positive integers, with exactly Bennett’s list of exceptions. The coprime part was already proven by Styer (2024). The non-coprime exceptions were not found in the literature:

a	b	c	Solutions (x, y)
6	2	4	(1, 1) and (2, 5)
15	6	9	(1, 1) and (2, 3)
6	3	1215	(4, 4) and (5, 8)
280	5	275	(1, 1) and (2, 7)

Steiner’s circuits. For every odd $q \geq 3$ and every length $n \leq 2500$, the only circuits of $qx + 1$ are the families $q = 2^n \pm 1$ and the known circuits of $3x + 1$ and $5x + 1$. Outside the families, only $q = 5$ has a positive circuit.

5.2 Exact closure

A chain criterion for k/p. Let $k \geq 4$ and p a prime not dividing k . Then $k/p = 1/x + 1/(xY) + 1/(xZ)$ has a solution exactly when $p + 1$ has a divisor d with $d \equiv -p \pmod k$. For $5/p$, such a closure always exists when $p \not\equiv 1 \pmod 5$.

A step criterion for Erdős-Straus. For primes $p \equiv 1 \pmod{24}$, $4/p$ has a closure with $x = (p + 7)/4$ exactly when p or $(p + 7)/4$ has a prime factor $\equiv 3, 5 \pmod 6$. This was recomputed on all 3202 such primes below 300,000, with no exception.

5.3 Prime windings

Sierpiński problems in negative bases. In base -8 , the smallest non-trivial k with $k(-8)^n + 1$ composite for all $n \geq 1$ is 246, with covering set $\{5, 7, 13\}$. In base -6 it is 1886. The hardest case below it, $k = 154$, needed a certified prime of 3,312 digits, at $n = 4253$.

Balanced ternary reads class numbers. Let $p \equiv 3 \pmod 4$ be a prime for which 3 generates the squares modulo p . Then the sum of the trits $(+1, 0, -1)$ over one period of $1/p$ equals $h(-p) \cdot (1 - (2|p))$, where $h(-p)$ is the class number of $\mathbb{Q}(\sqrt{-p})$. This is zero for $p \equiv 7 \pmod 8$ and $2h(-p)$ for $p \equiv 3 \pmod 8$. It was recomputed on 497 primes, with no exception. It is a variant of Girstmair’s work, by his method. The trit register of the theory thus carries a deep invariant of arithmetic.

5.4 Scale and orbits

Gleason polynomials. The polynomial whose roots are the centres of period n of $z^2 + c$ has discriminant $\equiv 1 \pmod 8$ for every n . Its Galois group is the full symmetric group for $3 \leq n \leq 11$.

Lonely runners in arithmetic progression. Take speeds $a, a + d, \dots, a + (n - 1)d$ with $\gcd(a, d) = 1$, and put $N = 2a + (n - 1)d$. The loneliness is exactly $1/2$ for even d and $1/2 - (n - 1)/(2N)$ for odd d . This was recomputed on 116 progressions.

Kaprekar’s map. The four-digit Kaprekar map is fully classified in every even base $2^e \cdot m$. This proves a remark left open in the literature (COST, Remark 28).

5.5 One pattern across families

Two patterns recur.

- **Closure sits on best approximations.** For $ax + 1$, every cycle obeys $a^k < 2^n \leq (a + 1)^k$. Each pair (n, k) therefore admits at most one multiplier a .
- **Cycles appear where doubling closes.** In the Kaprekar, Ducci and Chebyshev maps, cycles and long transients appear exactly where m divides $2^j \mp 1$.

Both are statements about the rest between two windings. That is where the closure reading pointed the search.

6. Evaluation

6.1 What the judges found

The judges did what the occasion said a judge must do. They found real errors.

- **A gap in the Pillai computation.** The code took the first bound that held and never checked larger values. For 139 base pairs the searched range was too small. A continued-fraction routine also used numerators where it needed denominators. The judge recomputed the bound exactly, searched to 10^{250} , and confirmed the theorem. Without that check, the result would have rested on a flawed computation.
- **A refutation.** One rule in the exact-closure family was false. The counterexample $p = 29$, $a = 7$, $x = 9$ shows that $7/261$ is not a sum of two unit fractions. This was recomputed independently.
- **Known results.** Sixteen claims were already in the literature, among them Styer (2024) on Pillai, Atkin (1966) on Collatz's permutation and Franco (1990) on two-step cycles.

6.2 What changed between the rounds

The machinery was the same in both rounds: the same agents, the same tools, the same time. One thing differed. Round one asked where closure breaks. Round two asked what every closure obeys.

Round one produced nothing. Round two produced 61 verified theorems. In this experiment, steering by closure turned an empty search into a productive one.

This is what the closure rule predicts. A search for exceptions to a true statement finds none. A search for the structure of what closes finds structure, and structure can be proven.

6.3 What it is not

No famous problem was solved. Collatz, Erdős–Straus, the Lonely Runner, Bennett in general and the Riemann hypothesis remain open. Several of the new theorems are short corollaries of known results. Several are finite in range. “New” means not found, not certified absent.

The rule steered where to look. It did not supply the decisive argument for any famous problem. That argument, valid for all cases at once, is the closure Poincaré described. The probes did not produce it.

The distinction matches the one the formal theory itself keeps: proven, computed, corresponded, open.

7. Open threads

Three threads came closest to something larger.

1. **Erdős–Straus, hypothesis H1.** Every prime has a closure $4/p = 1/x + 1/y + 1/z$ in which the smallest denominator divides another. Checked to 10^7 . A proof would settle Erdős–Straus.

2. **Two odd steps.** Only $q = 5$ and $q = 181$ give a cycle of $qx + 1$ with exactly two odd steps. Checked to $n = 2000$. A cycle of this kind means $q^2 + D = 2^n$ with D dividing $2^t - 1$. Settling all t for all n would settle the question.
3. **Trit windows of $1/p$.** The single-trit sum of $1/p$ reads the class number. Whether the count of closed trit windows over a period of $1/p$ has an exact formula is open. The formal theory's closure lemma, $P(\text{close}) = (1/3)[1 + \hat{p}(\omega)^3 + \hat{p}(\omega^2)^3]$, gives the baseline to compare against.

Annotated references

Knight, W. (2026). "AI's mathematical prowess was just the beginning." *WIRED AI Lab newsletter*, 14 October 2026. *Why read?* The occasion. It reports the MIT study, the Meta-Duke transmission result and the Harvard-NTT threshold in one place. Reading advice: the first item and the two "Elsewhere on the Frontier" items.

Manso, G. (2026). "A Benchmark for Scientific Reasoning: Frontier AI on Unseen Science." *OpenAI Forum*, 21 July 2026. <https://forum.openai.com/public/videos/event-replay-a-benchmark-for-scientific-reasoning-frontier-ai-on-unseen-science-2026-07-21> *Why read?* The only public account of the MIT study's design: screener, respondent, judge, premises, answer keys and traps. Reading advice: the transcript, minutes 2 to 9.

Hu, Y., Ghahremani, P., Aghajani, R., Pasupuleti, M. and Bhattacharjee, H. (2026). "Reward Hacking as a Transmissible Strategy in Multi-Agent Coding Systems." *arXiv:2610.12822*. <https://arxiv.org/abs/2610.12822> *Why read?* It shows that a trick learned against a test spreads between agents, and proves the threshold $R = k\theta_1/u_0$. The reason every judge in this experiment wrote its own code. Reading advice: Sections 3, 5.1 and Theorem 5.1.

Crawley, E. and Tanaka, H. (2026). "Ecology of AI Agents: Collaboration Creates a Population Threshold for Takeoff." *arXiv:2610.12436*. <https://arxiv.org/abs/2610.12436> *Why read?* A threshold that exists only through cooperation, in closed form, with a worked example. A theoretical model, not a measurement. Reading advice: Sections 2.2 and 3.1.

Konstapel, J. (2026a). "Vacuum.Net Theory: hoe massa, energie en ruimte één net vormen." *constable.blog*, 28 September 2026. <https://constable.blog/2026/09/28/the-vacuum-net-theorycomplete-derivation/> *Why read?* The theory in plain language: strand, winding, knot, mesh, and the two roads of energy.

Konstapel, J. (2026b). "The Mathematics of Closure: A Map of Open Problems in the Vacuum.Net." *Version 1.0*, 15 October 2026. *Why read?* The map of 25 open problems in eight closure families from which the probe families were drawn.

Konstapel, J. (2026c). "The Vacuum.Net Theory: A Formal Statement." *Version 3*, 16 October 2026. *Why read?* Five axioms, nine definitions and eleven proven theorems, including the 9-of-27 closure of trit windows and the closure lemma. Reading advice: Sections 1, 2 and 4.

Poincaré, H. (1908). *Science et méthode*. Paris: Flammarion. **Book I, Chapter III, "L'invention mathématique."** *Why read?* The first-hand account of insight arriving whole: the bus at Coutances. A mathematician describing what he lived through, not a theory of it. Reading advice: Chapter III alone.

Hadamard, J. (1945). *The Psychology of Invention in the Mathematical Field*. Princeton University Press. *Why read?* Accounts collected from working mathematicians, Einstein among them, showing the same order: the

whole first, verification after. Reading advice: Chapters II to IV.

Simons, J. and de Weger, B. (2005). "Theoretical and computational bounds for m-cycles of the $3n + 1$ problem." *Acta Arithmetica* 117, 51-70. Why read? Collatz cycles excluded through the convergents of $\log_2 3$. The published form of the closure-and-rest reading for $a = 3$.

Steiner, R. P. (1977). "A theorem on the Syracuse problem." *Proceedings of the 7th Manitoba Conference on Numerical Mathematics*, 553-559. Why read? The origin of circuits for $3x + 1$, which the probes extended to all odd q up to length 2500.

Lagarias, J. C., ed. (2010). *The Ultimate Challenge: The $3x + 1$ Problem*. American Mathematical Society. Why read? The standard survey, with an annotated bibliography. The place to check whether a cycle result is known.

Mihăilescu, P. (2004). "Primary cyclotomic units and a proof of Catalan's conjecture." *Journal für die reine und angewandte Mathematik* 572, 167-195. Why read? It settles the case where the rest between two powers is exactly 1. Reading advice: the introduction states the result in one page.

Bennett, M. A. (2001). "On some exponential equations of S. S. Pillai." *Canadian Journal of Mathematics* 53, 897-922. Why read? The theorem that $a^x - b^y = c$ has at most two solutions, and the conjectured list of exceptions that the probes confirmed up to 300.

Styer, R. (2024). arXiv:2402.04428. <https://arxiv.org/abs/2402.04428> Why read? The proof of Bennett's conjecture for coprime bases up to 3600. It showed that part of a probe result was already known.

Elsholtz, C. and Tao, T. (2013). "Counting the number of solutions to the Erdős-Straus equation on unit fractions." *Journal of the Australian Mathematical Society* 94, 50-105. Why read? The Type I and Type II parametrisation of all closures of $4/p$, against which every Erdős-Straus probe was checked.

Girstmair, K. (1994). On the digits of $1/p$ and class numbers. *Acta Arithmetica* 67. <https://bibliotekanauki.pl/articles/1391596.pdf> Why read? The method behind the balanced-ternary class-number result.

Kravitz, N. (2021). "Barely lonely runners and very lonely runners." *Combinatorial Theory* 1. Why read? The spectrum conjecture for lonely runners, proven by the probes on one new family.

Laurent, M. (2008). "Linear forms in two logarithms and interpolation determinants II." *Acta Arithmetica* 133, 325-348. Why read? The bound that the judge used to repair the Pillai computation. Reading advice: the main theorem only.