The Mathematics of Grego's Limbo Match

and how merit is weighed over luck

Three studies of a memory, strategy and lottery game

GREGO.HU/LIMBO

a companion to the game — playing explains everything

by Grego (Gergely Földvári)

Collected edition, 2 September 2026. The three studies were written and revised through 2026.

Grego's Limbo Match is a memory–strategy–lottery game. P blue chips numbered 1 to P sit face-down in a middle stack; their matching reds lie face-down around a dial. A player flips the top blue, then searches the reds for its pair within a flip limit — or gambles the clover for a blind guess. A missed search sends the blue to the bottom of the stack to resurface last with a better chance: the Limbo. These three studies ask what the game is worth mathematically: how many categories the Feast admits, how two perfect readers of one shuffle differ, and where the depth of the game actually comes from. The recurring question is the one in the subtitle — how a scoring system can weigh merit above luck, and be shown to do so.

 DOI 10.5281/zenodo.22260942
grego.hu/lm · grego.hu/lm.pdf · grego.hu/limbo

Contents

A NOTE ON PART III

Part III was written by Claude (Anthropic) at the author's direction, from the shipped rules and from simulations of them. Its findings were checked, and in several places corrected, by the author; the document's own closing section records where that happened, including one claim that was written, tested and thrown away. It is included here as commissioned analysis, not as independent review, and the responsibility for it is the author's.

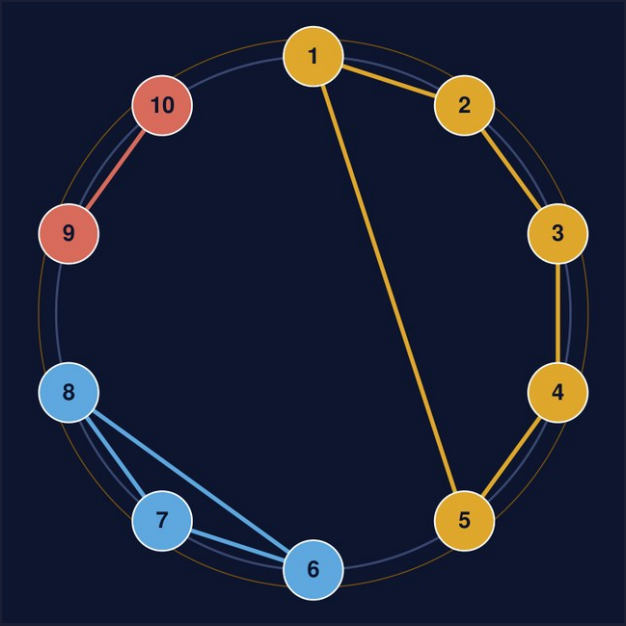
SEQUENCES

A389262 — the Collective Survival Threshold, contributed by the author: the least number of choices that keeps the hundred-prisoners loop strategy above one half. A000041 — the partition numbers, which count the shapes a shuffle can take.

GREGO'S LIMBO MATCH™

The Maths of the Feast

The feasible tally, the worth of a category, and the merit that ranks a champion



one shuffle of P = 10 · loops {5, 3, 2} · a partition of 10

grego.hu/limbo · OEIS A389262 (CST)

The Feast · 220,407 categories

The object, and what a category is

Grego's Limbo Match is a memory–strategy–lottery game. P blue chips numbered 1…P sit face-down in a middle stack; their P matching red chips lie face-down around a marked dial. The player flips the top blue to read its value, then searches the reds for its pair within a flip limit, or gambles the clover for a blind guess and banks points for unspent flips. A missed search sends the blue to the bottom of the stack to resurface last with a better chance: the Limbo. Jackpot Jack, the score ceiling of a category, is P×(F+1).

A category fixes the setup with four dials

P — pairs (2 to 21). F — the flip limit per search (1 to P). L — the limbo come-back chances (0 to P). LL — the loop-length window [MIN, MAX] of the shuffle. A shuffle lays the reds in a secret order; it falls into loops, and the set of loop lengths is the whole story that the pages below count and weigh.

The F = P edge

When F = P the flip limit already lets every red be revealed, so a miss is possible only by re-flipping a red, which is non-virgin and denies the limbo. Limbo fires only when F < P. At F = P the L dial is inert, so an F = P category exists only at L = 0.

Dial	Meaning	Range	Note
P	pairs — one blue + one red	2 … 21	board size
F	flip limit per search	1 … P	F = P makes every red reachable
L	limbo come-back chances	0 … P	fixed at 0 when F = P
LL	loop-length window [MIN,MAX]	1 … P	selects the shuffle pool

The dial face also carries the clover (a blind-guess lottery for the full flip bonus), the pin (rotate all reds until a chosen red sits on its own dial number) and the eye (peek every red, at the cost of the record). These are strategy, not category axes.

The canon — the World Record line

The World Record is one fixed category per board. On every WR rung the limbo count is L = 1 and the loop range is MIXED (any length); the flip limit is the CST, A389262(P) — the Critical Survival Threshold, the least flip-cap at which more than half of all P-shuffles are winnable under the 100-prisoners loop strategy (Golomb & Gaal, 1998). Each WR rung sits on the fair, better-than-even line. **The WR line runs P = 3 to 21 — nineteen rungs. At P = 2 the CST equals P, so P = 2 carries no WR rung; every P = 2 category is a Custom Record (F = 1 with any L, or F = 2 with L = 0). The world shuffle may still deal a P = 2 board.**

P	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
CST	2	3	3	4	5	5	6	6	7	8	8	9	9	10	11	11	12	12	13

WR flip limit per rung — the CST sequence A389262, by Grego (Gergely Földvári).

The field — Custom Records

Custom Records open the whole field around that calibrated centre. A player cooks a category by choosing F (1 to P), L (0 to P) and a loop window LL = [MIN, MAX]. The nineteen WR rungs are the fair midline; CR is everything else reachable from it.

The maths — shuffles, loops, partitions

A shuffle of P reds is a permutation, and its loops are its cycles. A loop pattern is a partition of P — a way to write P as a sum of parts (for P = 10, one pattern is 5 + 3 + 2). The number of shuffles sharing one loop pattern is the cycle-type count; summed over every partition of P these counts equal P!.

A category's shuffle pool

The window [MIN, MAX] gathers every partition whose parts all fall inside it; the pool size is the sum of those partitions' shuffle counts. At P = 10 the window [1, 4] gathers 23 partitions (632,736 shuffles) and [2, 5] gathers 7 (345,681 shuffles). The window [1, k] — all loops at most k — is where the CST lives: the smallest k whose pool passes P!/2. These pools are large: a full P = 13 board has 13! = 6,227,020,800 shuffles, and every category's pool is an exact sub-sum of that total, so a score is placed against a field whose size is known to the last permutation.

Distinct pools

Redundant bounds collapse. At P = 10 with MIN = 3, raising MAX from 7 to 8 or 9 selects the same partitions (7 is the longest loop that fits), so 3–7, 3–8 and 3–9 are one category; 3–10 adds the single 10-loop and stays distinct. A window that no shuffle can satisfy is empty and is not a category. The distinct, non-empty pools per P are the LL categories column below.

The count — 220,407 categories

For each P the categories multiply out as Feast(P) = (distinct LL categories) × (valid F, L pairs). F runs 1 to P and L runs 0 to P, and F = P forces L = 0, which leaves P^2 valid (F, L) pairs per board:

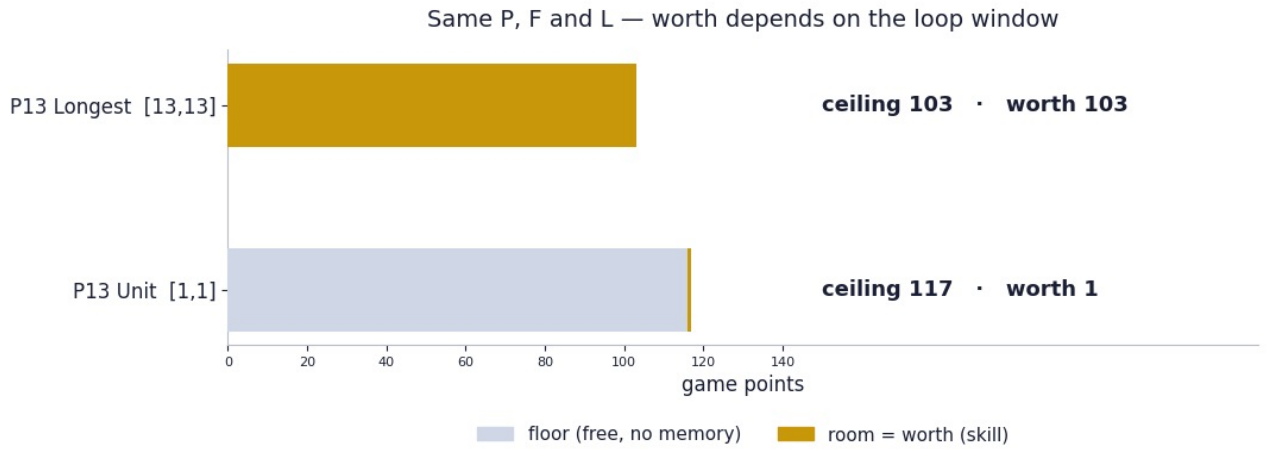
P	partitions	LL cats	F,L pairs (P^2)	Feast(P)
2	2	3	4	12
3	3	4	9	36
4	5	7	16	112
5	7	8	25	200
6	11	13	36	468
7	15	14	49	686
8	22	19	64	1,216
9	30	22	81	1,782
10	42	28	100	2,800
11	56	30	121	3,630
12	77	40	144	5,760
13	101	41	169	6,929
14	135	49	196	9,604
15	176	55	225	12,375
16	231	63	256	16,128
17	297	67	289	19,363
18	385	80	324	25,920
19	490	83	361	29,963
20	627	95	400	38,000
21	792	103	441	45,423
2–21				220,407

LL categories are distinct shuffle pools after collapsing redundant bounds; every column of partition counts sums to P!.

Summed over P = 2 to 21 the feasible tally is 220,407 categories: 19 World Records (the MIXED slice with F = CST and L = 1) and 220,388 Custom Records. Weighted by the shuffle counts above, the Feast is the ground the ranking stands on — every score placed against a category whose size and difficulty are known exactly.

Worth — difficulty in the game's own currency

A category's worth is the most merit any deal inside it can ever hold, on the same 1…441 scale as the merit itself. It answers one question: how much does knowing the shuffle buy here? Worth is window-aware. A 13-pair Unit board and a 13-pair Longest board share the same P, F and L, yet their worth is 1 and 103.



Worth is the room a shuffle leaves for memory, maximised over the shapes the window allows. A Unit board hands everything over free (room 1);

a Longest board hides a full trail (room 103).

How worth is computed

- For every loop shape the window allows, worth takes room = max(1, ceiling floor) and keeps the largest. The ceiling depends only on the shape, never on the blue order, so worth is exactly computable by sweeping the shapes — at most 792 at P = 21. Worth orders the Champions' Table house, the empties sort, GLSX prestige and GLTX difficulty.

Floor and ceiling — the two honest bounds

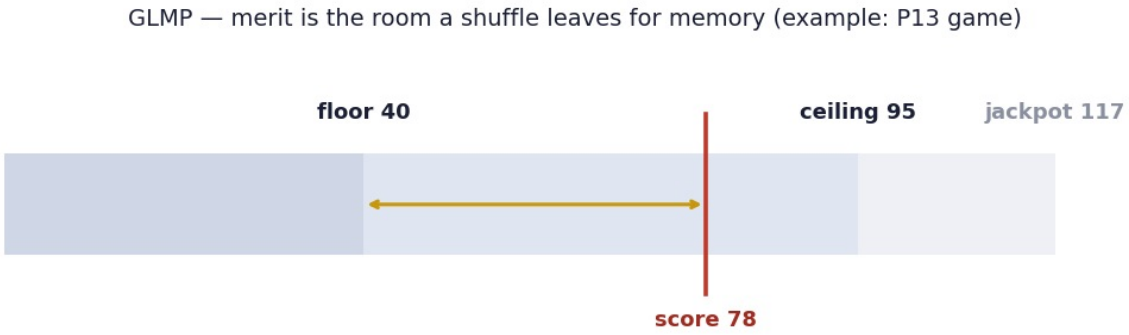
Merit is measured against the actual shuffle the nick played, between two bounds that both err on the safe side.

The floor — what the shuffle hands over free

The floor is the no-memory trail-follower. To find red v it walks the trail from dial position v, which takes exactly the - length of v's loop; a match on flip k scores 1 + (F k), and a trail longer than F is never found. The floor depends only on the shuffle and F. It is the luck the board gives everyone.

The ceiling — the most a perfect memory can take, without gambling

The clover replaces flipping the blue, so a full 1 + F needs knowing both blue and red; a known red otherwise costs one tap for F. Limboed blues return in order, yellow-ringed and fanned, so they are all knowable; trail-following unseen reds is normal play, and the last unseen red is free by elimination. The deliberate-limbo harvest lets a spare slot force a known blue down to be clovered on return for 1 + F instead of F, at the cost of carrying its value and position across the wait. The ceiling takes the best of these perfect-memory lines. The virgin rule bans only re-flips; burning matched or known reds is legal.

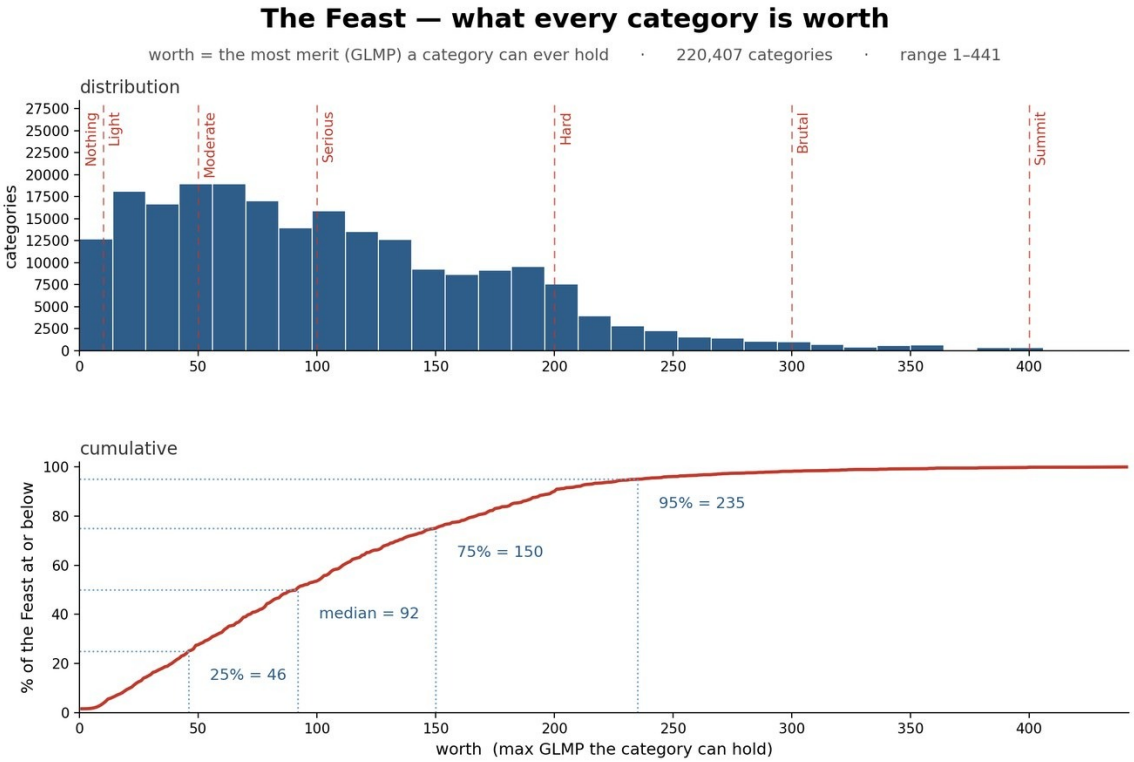


GLMP = min(S, ceiling) − floor = min(78, 95) − 40 = 38

Between the free floor and the risk-free ceiling lies the room a memory can convert into points; capped at the score, that room is the merit.

The Feast, by worth

Weighted by their shuffle counts, the categories spread across the worth scale as below. The mass sits in the Serious band (100–199) at 36% of the Feast; the Summit (400–441) is razor-thin at 506 categories, reachable only where a long-loop window and a high L let a perfect memory harvest deliberate limbos. Worth and board size are independent axes — most low-worth categories still carry ten or more pairs.



Distribution and cumulative worth across all 220,407 categories.

statistic	value
mean	104.5
25th percentile	46
median	92
75th percentile	150
90th percentile	200
95th percentile	235
99th percentile	337

worth	name	categories	share
1	Nothing to prove	3,315	1.5%
2–9	Trivial	3,366	1.5%
10–49	Light	53,587	24.3%
50–99	Moderate	57,214	26.0%
100–199	Serious	80,165	36.4%
200–299	Hard	18,818	8.5%
300–399	Brutal	3,436	1.6%
400–441	Summit	506	0.2%

Left: worth percentiles across the Feast. Right: the eight worth bands.

GLMP, GLMR and GLM100 — merit and its ranking

GLMP — the merit of a game

GLMP = max(1, min(S, ceiling) − floor). It reads the raw game points S against the shuffle's own floor and ceiling — nothing else. The floor of 1 is the point for understanding and playing. Across the Feast GLMP runs 1 … 441, the same scale as worth, because worth is the largest GLMP a category can hold. On a P = 13 Longest board every trail is length 13 — longer than the flip cap — so a no-memory player finds nothing and the floor is 0, while a perfect memory reaches a ceiling of 103; a nick who scores 90 there earns GLMP 90. On a P = 13 Unit board (floor 116, ceiling 117) the same nick might score 116 and earn GLMP 1, because the board asks nothing.

The blind run

One case needs its own bounds. Before the first flip no red has been revealed, so the leading run of clover hits is **provably** uninformed — the log shows it, nothing is inferred. Such a player cannot use the board, and the board does not help them: an all-blind run comes off at 1 in P! whatever the shuffle, measured identically from floor 0 to floor 22. Charging them the trail-follower's floor would deduct points that were never free to them, and the memory ceiling would measure a route they never took.

The only skill in a blind run is **elimination**. Matched reds stay face-down and look exactly like unmatched ones, so remembering which are spent is what turns a 1/P guess into 1/(P − j) — worth 26 times at P = 5 and 114 million times at P = 21. So the run is measured against blind bounds, **floor = (1 + F)·n/P** for no memory and **ceiling = (1 + F)·Σ 1/(P − j)** for perfect elimination, and the usual clamp applies: scoring above what perfect elimination expects is luck and earns nothing. From the first flip the game continues as a genuine (P − n)-pair board and is scored normally.

A blind jackpot is therefore worth 2 merit at P = 3, 4 at P = 5, 20 at P = 13 and 37 at P = 21 — small, scaling with the board, and far below the hundred-odd that reading a shuffle earns. **A lucky player may hold a high seat and still rank low**, which is the honest answer to anyone who fears it is pointless to think against a lottery.

GLMR — where a nick stands

A nick is ranked by its best single GLMP. GLMR = 1 + (the number of nicks with a strictly higher GLMP): a peer-versus-peer ladder with honest joint ranks and no tiebreak, so the best players sit at the top rather than against the 441 ceiling.

GLM100 — the certified hundred

The GLM100 is the head of the ladder: the nicks at GLMR 1 to 100. Matched reds stay face-down, visually identical to unmatched ones, so remembering which red is already spent is provable skill that even a two-pair game asks — the record score itself encodes merit. Within a single category the record is the best score, faster time breaking a tie.

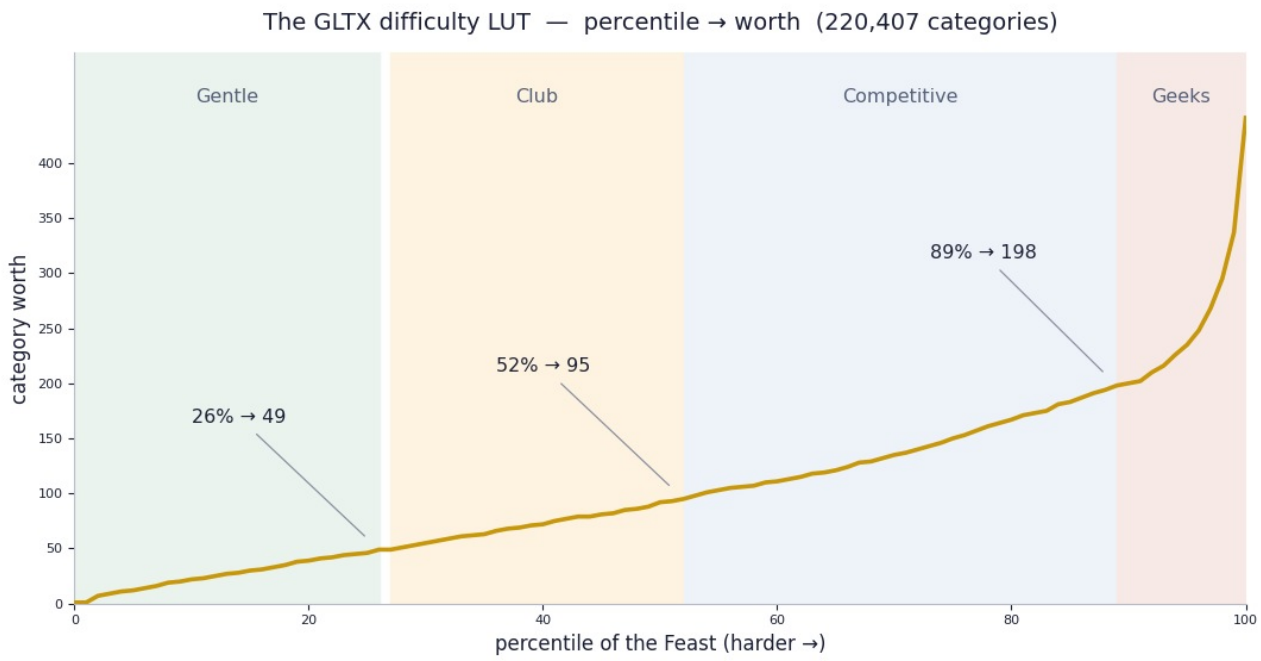
Where worth and merit are used

The Champions' Table

The house is ordered by worth — lighter categories first — and the empty categories are sorted the same way. Each category holds one record, its best score with faster time breaking a tie; across the house the champions are ranked by GLMP then GLMR, with the GLM100 as the certified head.

GLTX — tournaments buy difficulty by percentile

An organiser sets difficulty as a percentile band, and the slider reads the LUT below, mapping 0–100 to worth so “harder than 75% of the game” is plain. The presets are Gentle (worth 1–49), Club (50–95), Competitive (95–198) and Geeks (198–441).



fi

The GLTX difficulty LUT — percentile of the Feast category worth.

GLSX — sponsorship price rests on worth

A category's sponsorship price combines its prestige — its worth, 1…441 — with its share, how busy it is against an average category, floored at a small per-category-day minimum.

The summa — the numbers this rests on

summa	value
Feast — total categories	220,407
World Record rungs	19
Custom Record categories	220,388
Worth and GLMP range	1 … 441
Median worth	92
Summit categories (400–441)	506
Jackpot ceiling	J = P × (F + 1)
Shuffles at P = 13	13! = 6,227,020,800

The nineteen World Record rungs

Each rung fixes L = 1 and the MIXED window; F is the CST. Jackpot Jack is J = P × (F + 1), and the worth is the most merit the rung can hold.

P	F	J	Worth	P	F	J	Worth
3	2	9	8	13	8	117	103
4	3	16	14	14	9	140	125
5	3	20	16	15	9	150	133
6	4	30	25	16	10	176	158
7	5	42	36	17	11	204	185
8	5	48	41	18	11	216	195
9	6	63	55	19	12	247	225
10	6	70	60	20	12	260	236
11	7	88	77	21	13	294	269
12	8	108	96				

P = 2 carries no rung: its CST equals P.

The tools on the dial

Clover— a blind-guess lottery: tap any red for 1 point plus the full flip bonus when it matches. **Pin**— rotate all reds, keeping the shuffle, until a chosen red sits on its own dial number. **Eye**— reveal every red face-up; while used, the game does not qualify for a record. **Limbo and the virgin rule**— a missed search sends the blue to the bottom to return last; the limbo is granted only when no red is re-flipped during the search.

The Feast is the ground the ranking stands on

Every score is placed against a category whose size and difficulty are known exactly: its shuffle pool counted to the last permutation, its worth swept over every loop shape, its merit measured between an honest floor and an honest ceiling. The Feast holds 220,407 categories. Worth alone is the lever. It tells a trivial board from a brutal one wherever the two share a size, and it lets a small board played perfectly outrank a large one shown up but not solved. A prodigy needs no age, and memory checks none.

References

A389262 (CST) — Grego (Gergely Földvári), 2025. Golomb & Gaal, On the Number of Permutations on n Objects with Greatest Cycle Length k , Adv. Appl. Math. 20 (1998). Wikipedia, 100 prisoners problem .

Glossary — every abbreviation in this paper

Collected from the text of this document. Every term below is used above; none is introduced here.

The board

term	meaning
P	Pairs. Blue chips 1…P in the middle stack, and their matching red chips around the dial.
F	Flip limit. Reds that may be turned in one search before the blue is missed.
L	Limbo allowance. How many missed blues may return at the tail, once each.
LL	Loop-length window [MIN, MAX]. Every loop in the shuffle lies inside it — no loop is required to equal either bound.
MIN / MAX	The two ends of the LL window.
MIXED	The full window, LL = [1, P]: any loop length may appear.
Jackpot	P × (F + 1). The most a category can pay, reachable only with luck.

The measures

term	meaning
floor	What the shuffle hands over free: the score of a no-memory trail-follower.
ceiling	The most a perfect memory can take WITHOUT gambling.
room	ceiling – floor. What is actually at stake in that deal.
worth	The largest room any deal in a category can offer, 1…441. A conservative lower bound.
GLMP	Merit of one game: max(1, min(S, ceiling) – floor). A provably blind run is measured against blind bounds instead — see The blind run.
GLMR	A nick's rank: 1 + the number of nicks with a strictly higher best single GLMP. Joint ranks, no tiebreak.
GLM100	The head of that ladder — the nicks at GLMR 1 to 100.

Records and services

term	meaning
WR	World Record. The 19 canonical rungs: MIXED, L = 1, F = CST.
CR	Custom Record. Any of the other 220,388 categories.
TR	Tournament Record, held inside one GLTX event.
GLRS	Grego Limbo Ranking System — the whole scoring and ranking model.
GLNX	Certified nicknames. A certified nick shows a green tick.
GLTX	Tournaments. An organiser buys difficulty as a percentile band.
GLSX	Sponsorship. A category's price rests on its worth and its share of play.

Numbers and sources

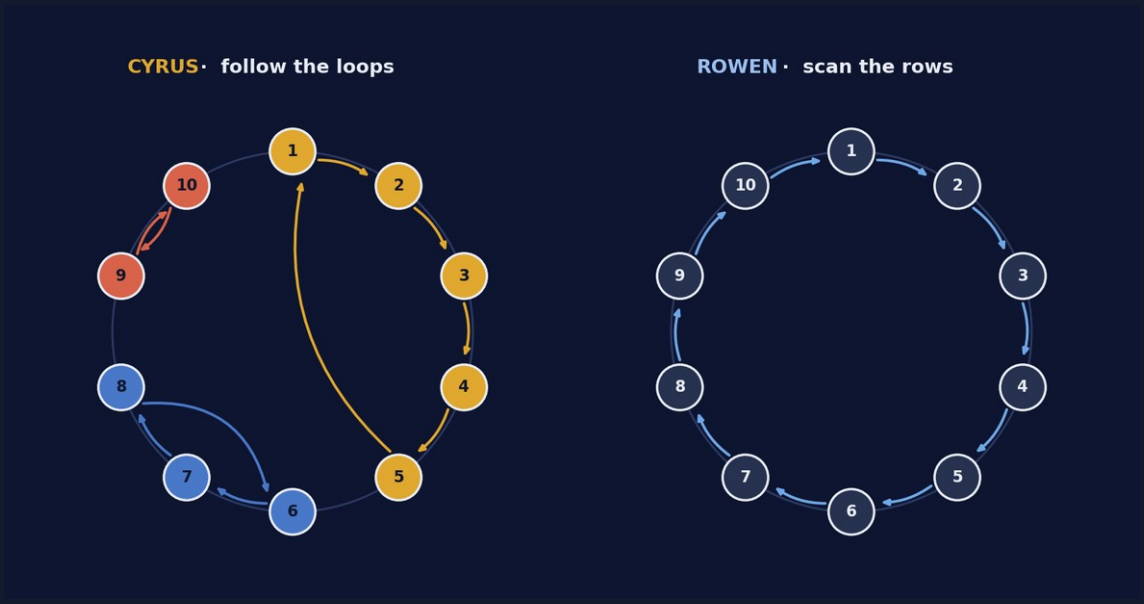
term	meaning
CST	Critical Survival Threshold — the WR flip limit per P, Grego's integer sequence.
A389262	The OEIS entry for the CST sequence, registered in Grego's name.
OEIS	The On-Line Encyclopedia of Integer Sequences.
Feast	All 220,407 categories together: 19 WR rungs and 220,388 CR.
LUT	Look-up table. Maps a 0–100 percentile to a category worth for the GLTX slider.

GREGO'S LIMBO MATCH™

MEMORY · STRATEGY · LOTTERY

The Two Readers

Cyrus the cycler, Rowen the rower — strategy, simulation,
and cognitive load across the Feast



Two readings of one Shuffle. Cyrus follows the loops — each chip he turns names the next chip to turn. Rowen scans the rows — positions in plain order. P = 10, loops {5, 3, 2}.

grego.hu/limbo · OEIS A389262 (CST) · revised 1 September 2026

THE QUESTION

Two readers of one Shuffle

A Shuffle hides P red chips in a secret order. Two players with perfect memory read it two different ways. Does the way they read change the score — and should the ranking care? **Cyrus** reads by loops . He flips the chip whose position equals the value he is hunting, reads what it shows, and flips that next — the chain leads itself, because in a Shuffle a position points to a value and that value names the next position. **Rowen** reads by rows . He turns the dial positions in plain order, 1, 2, 3, …, banking what he sees until the chip he needs turns up. Both remember everything; every flip costs time, and time enters the score. The arena is the **World Record line** — one category per P from 3 to 21, all **MIXED** (any loop length), **L = 1** limbo, and flip limit **F = CST** , the Critical Survival Threshold A389262(P). This paper settles the contest in three steps: a clean **tie** under perfect sighted play; a decisive **break** once the full toolset is in play; and a **field study** of how the gap moves across the Custom field. It closes with the human floor — what each style demands of a real memory as P grows — and what all of it means for the GLRS ranking.

THE GAME

Mechanics, in brief

A category fixes four dials: **P** pairs (2–21), **F** flip limit (1–P), **L** limbo come-backs (0–P), and the loop range **LL = [MIN, MAX]** . Flip a blue chip, search its red match inside F flips, or blind-guess with the clover and bank the unused flips as bonus. The score ceiling per category is the **Jackpot** :

$$\text{Jackpot} = P(F+1) \qquad \text{per category}$$

A missed blue is not lost: it drops to the bottom of the stack and resurfaces **last** — the **limbo** — with a far better second chance. One rule guards it: the **virgin rule** — flip a red twice in the first search and the limbo is void. The **World Record (WR)** line is the calibrated midline (MIXED, L = 1, F = CST); the **Custom Records (CR)** open the

Feast of 220,407	categories (the 19 WR plus 220,388 CR), counted and weighed in

the Feast . One structural fact drives everything below: at the CST, **F > P/2 for every P** , so a Shuffle can hold at most one loop longer than F.

P	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
F = CST	2	2	3	3	4	5	5	6	6	7	8	8	9	9	10	11	11	12	12	13

The WR flip limit per P — the CST sequence A389262, by Grego (Gergely Földvári).

THE FLOOR

Pure sighted play is a tie

Strip the game to plain matching — no clover, no tactics, both players sighted and flawless — and the two readers are **indistinguishable** . The reason is structural. Because F > P/2, a Shuffle holds **at most one loop longer than F** , so a sighted player suffers **at most one forced miss** per game — and the single L = 1 limbo always recovers it. Both therefore match all P pairs, every time. The contest collapses to a **flip race** :

$$\text{flips}_{\text{Cyrus}} = 2P - m\,(\#\text{loops}) \qquad \text{flips}_{\text{Rowen}} = 2P - r\,(\#\text{records})$$

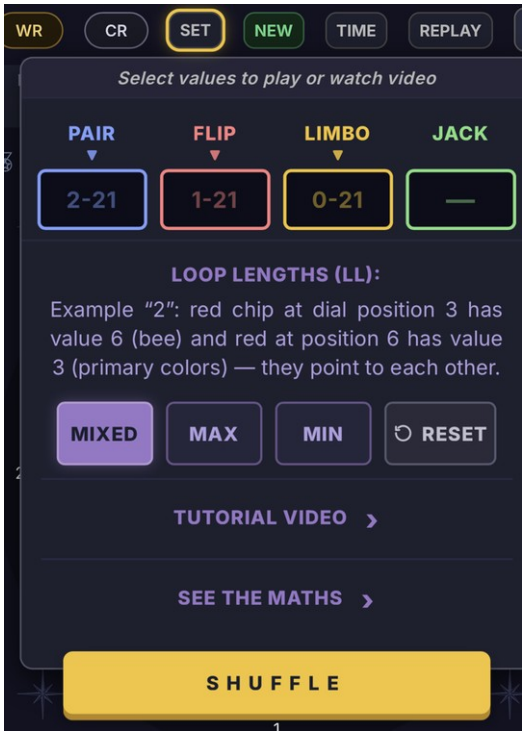
A loop is closed for free the moment its last unknown is forced (a record, for Rowen; a returning chain, for Cyrus). By the Foata–Schützenberger bijection the number of loops and the number of left-to-right records share

one distribution, with the same mean — the harmonic number:

$$\mathbb{E}[\,m\,] = \mathbb{E}[\,r\,] = H_P = \sum_{k=1}^{P} \frac{1}{k} \approx \ln P + \gamma$$

So their expected flips are **equal, category by category** . A Monte-Carlo party over all 19 WR categories (N = 80,000 each) confirms it to the decimal: **Cyrus 1961.50 · Rowen 1961.45 · difference +0.05.** The 95% confidence interval straddles zero; wins split 49.2% Cyrus / 48.9% Rowen. A coin-flip.

An athletes' relief — the flip race of Result A in stone: under flawless sighted play, the two readers finish level.



The live category picker on grego.hu — PAIR (P), FLIP (F) and LIMBO (L) set the three numeric dials; the loop window LL is chosen with MIXED / MAX / MIN. The worked example is a 2-loop: position 3 → value 6, position 6 → value 3 — the self-pointing chain Cyrus follows.

The table below is the exact engine of the tie: for each WR rung, the expected loop count equals the expected record count (H_P), so the expected flips (2P − H_P) match. The Jackpot is the per-category ceiling; the last column previews the full-game edge of the next section.

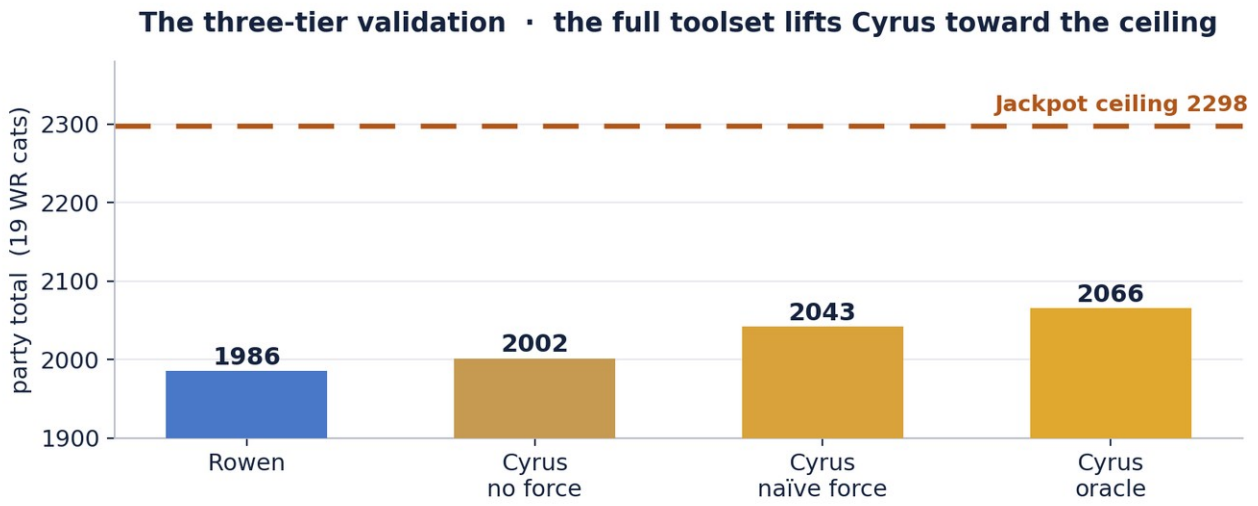
P	F = CST	Jackpot P(F+1)	E[loops]=E[records] = H_P	E[flips] = 2P − H_P	Cyrus edge (full game)
3	2	9	1.83	4.17	+1.33
4	3	16	2.08	5.92	+1.99
5	3	20	2.28	7.72	+1.87
6	4	30	2.45	9.55	+2.40
7	5	42	2.59	11.41	+2.99
8	5	48	2.72	13.28	+2.83
9	6	63	2.83	15.17	+3.43
10	6	70	2.93	17.07	+3.33
11	7	88	3.02	18.98	+3.87
12	8	108	3.10	20.90	+4.47
13	8	117	3.18	22.82	+4.34
14	9	140	3.25	24.75	+4.93
15	9	150	3.32	26.68	+4.89
16	10	176	3.38	28.62	+5.36
17	11	204	3.44	30.56	+5.95
18	11	216	3.50	32.50	+5.89
19	12	247	3.55	34.45	+6.46
20	12	260	3.60	36.40	+6.45
21	13	294	3.65	38.35	+6.96
Σ	—	2298	—	—	+79.7

Derived quantities computed exactly; the edge column is the measured full-game advantage (Result B). Jackpot sum 2298 is the party ceiling. **Reading.** The loop range adds no intrinsic difficulty on its own — perfect play ties regardless of how the Shuffle breaks into loops. Whatever separates the readers must come from the tools , not the bare match.

THE BREAK

The full game tips to Cyrus

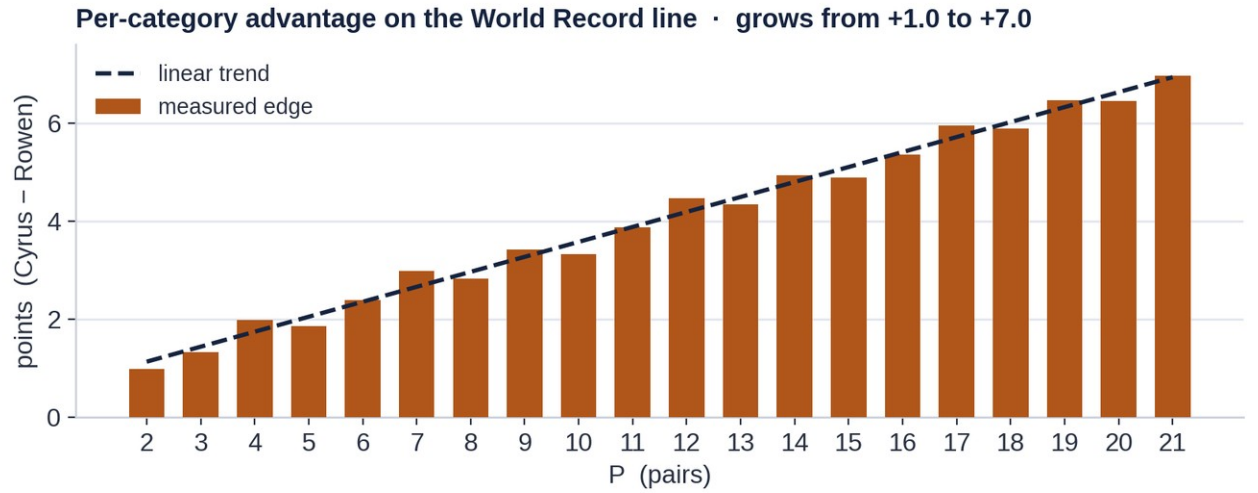
Three tools live above bare matching. The **clover** pays the maximum bonus (F + 1) for a blind match — better than any sighted search. The **limbo** returns a missed blue last, when the field is nearly known. And the **fanned tail** : as the stack empties, the surviving chips fan out, exposing by elimination what each must be. The decisive move is **force-limbo** , and only Cyrus can play it. Knowing the loops, he can decline the match on the longest loop's first blue — a clean miss, no second flip, so the virgin rule holds — and bank that blue into the limbo. Its long trace then converts into **two** tail bingos: the forced blue itself, and the fan-lit penultimate chip deduced once the fan reveals the limbo's value. Rowen scans blind; he cannot decline a match he has not yet located, so he cannot force. The result is a clean, growing separation:

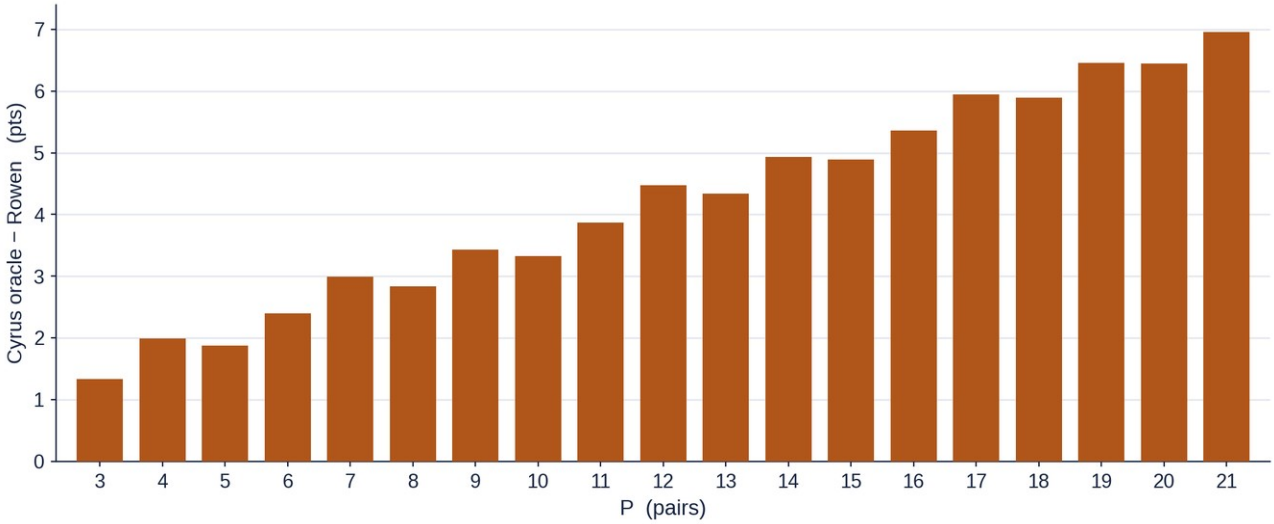


Party totals over the 19 WR categories (N = 40,000 each). Each tool Cyrus adds lifts him toward the Jackpot ceiling of 2298; Rowen sits at the sighted floor.

Player	Tools used	Party total	vs Rowen
Rowen	sighted + clover + limbo	1985.9	—
Cyrus — no force	+ cycle-lit fan	2001.7	+15.8
Cyrus — naïve force	+ force-limbo	2042.6	+56.7
Cyrus — oracle force	+ partition-aware stop	2065.6	+79.7
Jackpot ceiling	—	2298	—

The three-tier validation: no-force, naïve-force, oracle-force, each measured against Rowen on identical Shuffles. The full edge decomposes cleanly. Cycle-knowledge alone — just lighting the fan correctly — is worth **+15.8** . The force-limbo tactic adds **+63.9** , of which **+23.0** is the premium for choosing which loop to force well (oracle over naïve). And the advantage is not flat: it climbs steadily with P, from about +1.3 at P = 3 to +7.0 at P = 21.





Per-category advantage (Cyrus oracle − Rowen) by P. The edge rises almost linearly in P — more loops mean more structure for Cyrus to exploit.

THE TACTIC

Optimum stopping into the Feast

Force-limbo poses a real decision: which loop should carry the single limbo? Cyrus wants the **longest** loop — its trace seeds the richest tail — but he meets the loops one at a time as he traces, and must commit before he has seen them all. This is an **optimal-stopping problem** , a secretary problem played on cycle lengths, with one twist: if he waits too long, the natural forced miss (the loop beyond F) lands itself on whatever loop happens to overflow, possibly a short one. So the policy carries an **overflow hedge** — commit early enough to steer the limbo, late enough to catch a long loop. The fuel for that policy is the loop-length distribution — exactly the partition mathematics that sizes the Feast. Knowing how a random Shuffle of P tends to break (the cycle-type weights summing to P!) lets Cyrus set his stopping threshold optimally. That knowledge is the **+23.0** premium of oracle over naïve forcing: the same partition counts that price the categories in The Maths of the Feast also price this tactic.

$$\text{Feast} = \sum_{P=2}^{21} C_P \cdot P^2 = 220{,}407$$

P = pairs, 2 to 21 C_P = distinct loop-window categories at that P

P^2 = the valid (F, L) pairs: $P(P{+}1) - P$, since $F = P$ forces $L = 0$

THE FIELD

How the edge moves across CR

Beyond the WR midline, the Custom field exposes which dials the reader's edge actually rides on. Sweeping one dial at a time (anchored at P = 12) tells a consistent story: **Rowen barely moves** — he plays the sighted floor in every setting — while **Cyrus's score does the travelling** .

The loop window LL

Window	LL	Rowen ≈	Cyrus ≈	edge	edge %
Unit	[1, 1]	91.4	96.9	+5.5	6.1%
≤ Flip	[1, 8]	91.4	94.8	+3.4	3.7%
MIXED	[1, 12]	91.4	95.9	+4.5	4.9%
5-MIN	[5, 12]	91.4	96.5	+5.1	5.6%
Longest	[12, 12]	91.4	97.9	+6.5	7.2%

P = 12, F = 8, L = 1. The MIXED tie of the WR line is a knife-edge; constrain the window toward long loops and Cyrus pulls clear.

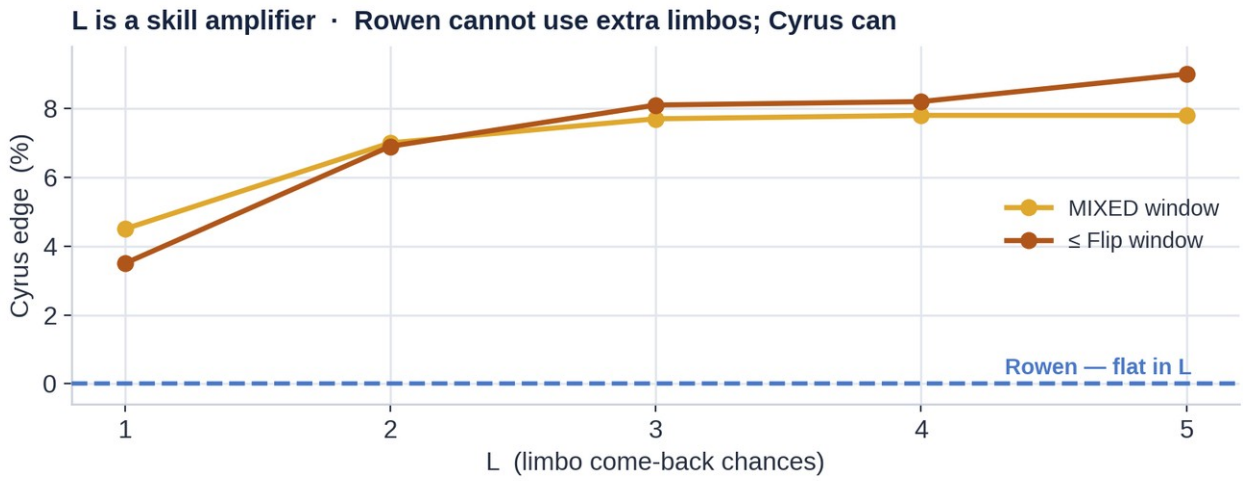
The flip limit F

F (MIXED)	edge	edge %
6	+3.4	4.9%
8	+4.5	4.9%
11	+6.9	5.5%

P = 12, L = 1. A larger flip budget gives Cyrus more room to convert structure into bonus; the edge grows with F.

The limbo count L — the sharpest axis

L is where the readers diverge most. With F > P/2 a sighted player has at most one natural miss, so **Rowen cannot use extra limbo slots** — his score is dead flat in L. Cyrus, who forces misses on purpose, fills them: each extra L lets him force another loop and deepen the fan tail. The edge roughly **doubles by L = 2** and plateaus near L ≈ 3, about the loop count H_P.



Cyrus edge versus L. Rowen is flat along the dashed line; Cyrus climbs until he runs out of loops to force.

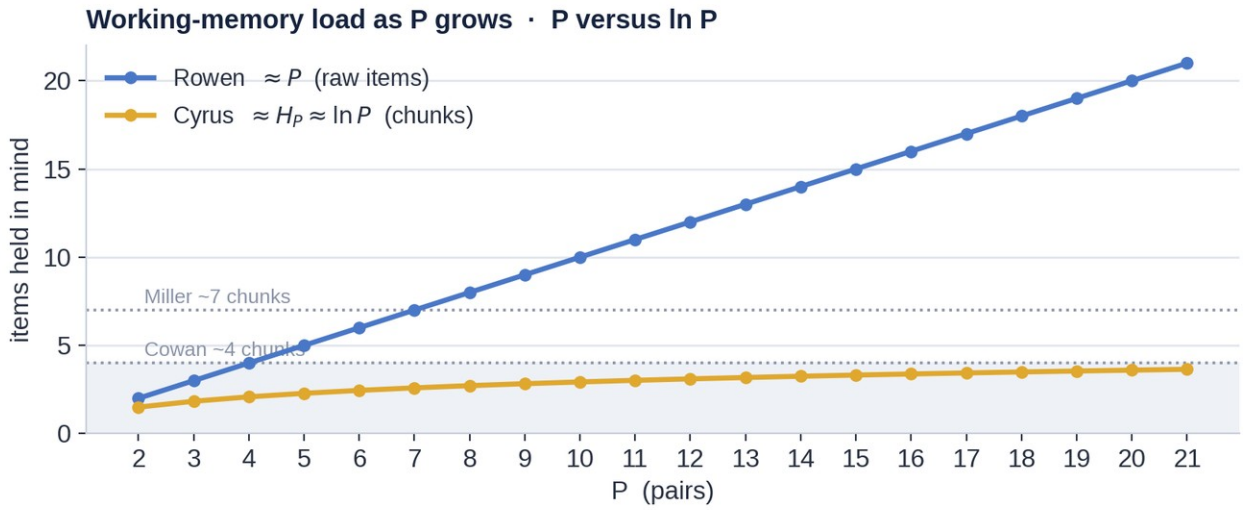
L	MIXED edge (pts)	≤ Flip edge (%)
1	+4.5	3.5
2	+7.0	6.9
3	+7.7	8.1
4	+7.8	8.2
5	+7.8	9.0

L is a pure skill amplifier for the loop-reader; F and the long-loop side of LL are the durable edge axes.

Memory, as P grows

The perfect-memory tie of Result A is the **floor** — the best possible case for Rowen, with his memory burden waved away for free. No human meets that assumption. Ask instead what each style actually demands of a real head. **Rowen carries an unstructured table.** Row-scanning offers no organizing principle: every red he turns is an arbitrary (position → value) fact he must bank, because any of them may answer a later blue. To find blue v he scans that mental table for where value v appeared — associative recall over a flat, structureless set that grows to size P. His load is **O(P) raw items** , like memorizing a random P-digit string. **Cyrus offloads onto the board.** Cycle-following is self-cueing : the value he just turned over names the next position to flip. During a trace he need not remember where to go — the chip points there. Tracing one loop costs **O(1) working memory regardless of its length** ; he holds only the start and the current link. What he tracks is just the set of loops , and a random Shuffle has only **m** ≈ **H_P** ≈ **ln P** of them — each a tidy, self-consistent chain held as a single chunk.

$$\text{load}_{\text{Rowen}} = \mathcal{O}(P) \qquad \text{load}_{\text{Cyrus}} = \mathcal{O}(H_P) = \mathcal{O}(\ln P)$$



Working-memory load against P. Rowen's raw-item count climbs past both the Cowan (~4) and Miller (~7) chunk ceilings; Cyrus stays inside the shaded band for every P the game allows.

P	Rowen items (≈ P)	Cyrus chunks (≈ H_P)	vs ceiling
4	4	2.1	Rowen at Cowan ~4
7	7	2.6	Rowen at Miller ~7
10	10	2.9	Rowen ~2.5× over Cowan
21	21	3.6	Rowen ~5× over Cowan

Human working memory holds ~4 chunks (Cowan) to ~7 (Miller). Rowen breaches the ceiling by P ≈ 4–7; Cyrus stays under it for all P (H_P reaches 4 only near P ≈ 31). The overload is not free — it couples straight into the score. A forgetful Rowen **re-flips** reds he has already seen: wasted flips (time, lost bonus), and worse, re-flipping a red in the first search **voids the limbo** by the virgin rule, costing whole pairs. Cyrus's self-cued trace almost never re-flips, so his limbo and bonuses survive. So the practical gap is ~0 at small P and **widens with P** — the same direction as the tool-driven edge. The two effects point the same way and compound.

THE UPDATE

What the scoring now does about it

This paper was written in May. Three rules have shipped since, and each settles a question it left open.

Merit can take a title

A seat no longer goes only to the higher score. A challenger also takes it with **equal score and faster time**, or with **higher merit at a time no slower**. Reading structure converts more of the room, and that now wins seats rather than arguments.

A blind run earns almost nothing

Rowen at least reads. A third player reads nothing and clovers blind. Before the first flip no red has been revealed, so that run is **provably** uninformed — a count of flips, not a guess at intent — and it is scored against blind bounds instead of the board's:

$$\textbf{blind floor} = (1 + F) \cdot n/P \cdot \textbf{blind ceiling} = (1 + F) \cdot \Sigma\, 1/(P - j),\ j < n$$

The board cannot help a blind guesser and no longer prices one: an all-blind run comes off at 1 in P! whatever the shuffle. What remains is **elimination** — matched reds stay face-down, so remembering which are spent turns 1/P into 1/(P − j). A blind jackpot is worth 4 merit at five pairs and 37 at twenty-one, against the hundred-odd that reading a shuffle earns.

A lucky player may hold a high seat and still rank low. That is the answer to anyone who suspects it is pointless to think against a lottery, and it is the sharpest form of this paper's result.

The Pin, and where it is refused

One tool arrived after this study. The **Pin** rotates the whole ring until a tapped red sits on its own number — composing the shuffle with a rotation, producing an entirely new loop structure, at no flip cost. It is the cycler's tool by nature, because its value is exactly what you already know.

It always creates a loop of length 1: the pinned red sits on its own number, which is the definition. So in a category whose window starts above 1 the Pin would manufacture the very thing that category was built to exclude, and the game refuses it there and flashes the window instead.

THE VERDICT

What the ranking should do

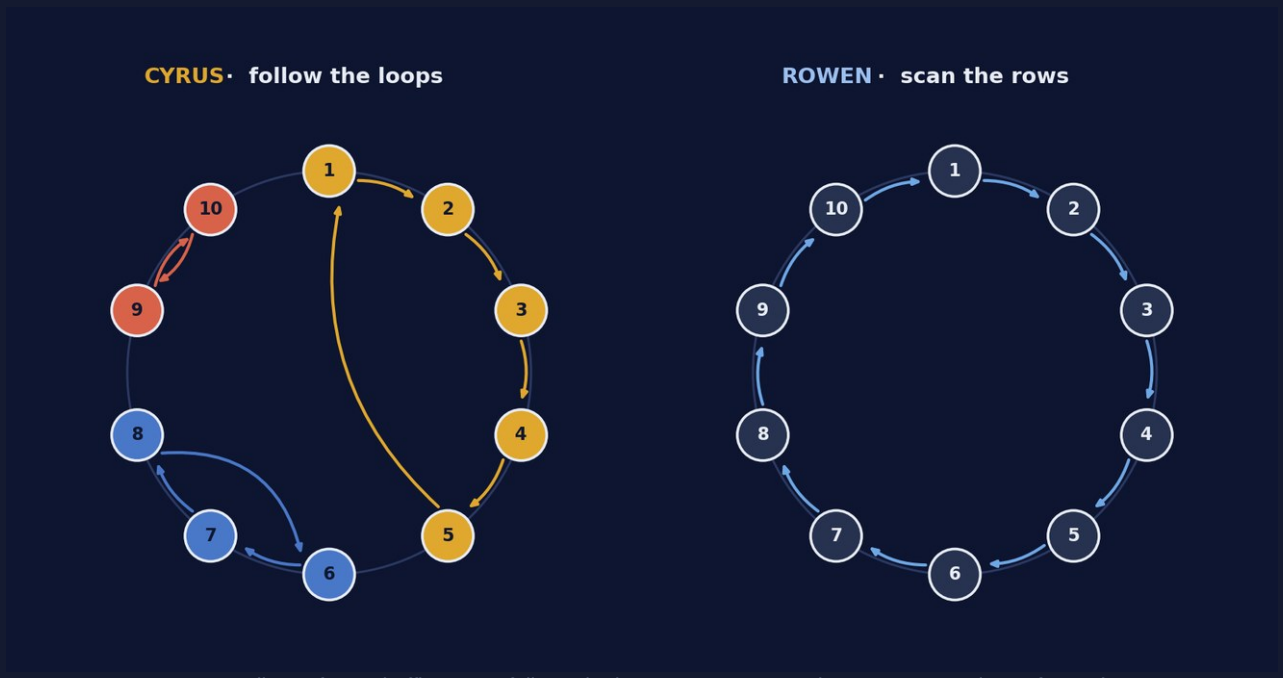
The findings converge on the **worth model** . Result A shows the loop range sets no intrinsic **flip-race** difficulty — under perfect play the readers tie regardless of how a Shuffle breaks. But the field study measures a second axis the flip race cannot see: the room between the no-memory floor and the perfect-memory ceiling, which the loop window genuinely widens. That room is the category's **worth** , and it is what the ranking weighs. No player is handed a loop-multiplier; the merit is **earned in play** , banked against the honest floor and ceiling — **GLMP** (merit points), ranked by **GLMR** and headed by the certified **GLM100** . The loop range earns its place twice: each window [MIN, MAX] is a **distinct Shuffle pool** with its own leaderboard — why the Feast counts them separately — and each carries its own **worth** , the room a memory can convert. The WR tie is revealed for what it is: a **knife-edge** , balanced only under flawless free play, that both the full toolset and real human memory break toward the reader who sees structure. Cyrus wins not because the loops are harder, but because he reads them.

In one line: LL sets the room; worth is what weights the categories, and the merit earned inside it is what the ranking rewards. Strategy and cognition do the rest — and both favour the cycler more sharply as P grows.

References. A389262 (CST) — Grego (Gergely Földvári), 2025. A000041 — partition numbers. A000142 — n!. Golomb & Gaal, On the Number of Permutations on n Objects with Greatest Cycle Length k , Adv. Appl. Math. 20 (1998). Foata & Schützenberger, fundamental bijection on permutations. Miller, The Magical Number Seven , Psychol. Rev. 63 (1956). Cowan, The magical number 4 in short-term memory , Behav. Brain Sci. 24 (2001). Monte-Carlo figures from the project simulations cyrus_rowen.py and full_game.py .

Across the Tiers

an A.I. analysis of
Grego’s Limbo Match™



Two readings of one shuffle. Cyrus follows the loops; Rowen scans the rows. P = 10, loops {5, 3, 2}.

One game, every age — where the depth comes from, and how deep it goes

Cycle mathematics · simulation of the shipped rules · revised 27 August 2026

Summary · the findings in one table

#	Finding	Where
1	Matched reds stay face-down, so visible difficulty is flat while real difficulty rises	§2
2	Every flip costs exactly one point — memory becomes economics	§4
3	A search costs exactly the length of the loop it walks	§5
4	Tracing does not raise the average — but lifts a clean sweep by ×24 at 8P and ×10,698 at 21P	§6
5	The record flip ladder sits on the Golomb–Dickman ratio, 0.62	§7
6	Perfect play clears the board about half the time at every rung from 5 to 21 pairs	§7
7	Memory is worth 21 points on an 8-pair board; every other skill together, about one	§8
8	Following the chain needs one number in mind, whatever the board size	§9
9	Clover beats a flip by exactly 2 ÷ k – 1 — always at one blue, never at three	§10
10	A sighted Pin rescues almost any dead board, and its power scales with knowledge	§12
11	Worth measures what must be earned, not what can be scored — 1 to 441	§15
12	Merit subtracts the shuffle's gift and caps the luck: GLMP	§16
13	A player is their best single game, which nobody else's play can erode	§17
13a	The limbo harvest: clovering a returning blue pays 1+F and is pure memory, worth +5 on one measured ceiling	§11
13b	Merit can take a title from a higher score — so a record score can go down, by design	§17
13c	In 26% of standing record pairs, the higher score carries the lower merit	§17
13d	Worth is a conservative lower bound, not an exact figure — a live record disproved the old claim	§15
14	No opponent is required, and one is always available — 220,407 standing records	§19
15	Three overlapping tiers, 3–7 · 3–13 · 2–21, with the rules identical in all of them	§20
16	The board moves to fit the player; the game is never handicapped to fit the age	§20

A last observation before the questions. Of the twenty findings above, only six — the eleventh, twelfth and the four added in this revision — required access to the shipped system. The rest follow from the rules alone, which means the game's depth is not a property of its implementation but of its design. That is worth stating plainly, because it is the difference between a clever program and a real game.

Revised 28 August 2026: a correction to section 15, a strategy the first edition missed in section 11, and the merit-based title rule in section 17. The correction was forced by a live record that exceeded a bound this document had published — see section 25.

1 · What this document answers

Grego made an observation while playing:

> *Even inside a medium category, the game offers an enormous depth of improvement in concentration, across all ages and abilities. To score higher or finish faster, a player can go deeper along a spectrum that feels almost infinite.*

This document tests that observation. It asks three questions. Is the depth real? Where does it come from? And can it be measured?

The short answer: the depth is real, it comes from one design choice, and it can be measured exactly. The measurements are in this document. Some of them are surprising even to a player who knows the game well.

Every number here was computed for this document — by exact combinatorics where an exact answer exists, and by simulation of the shipped rules where it does not.

But the phrase that matters most in Grego's sentence is **"across all ages and abilities"**, and that is what this document is named for. A game can be deep and still be the property of one kind of person — young, quick,

competitive, already good at it. The claim here is larger: that the *same* game, with the *same* rules and the *same* scoring, works for a five-year-old who cannot yet read, for an adult who wants a hard problem, and for an older player who wants to keep a sharp mind — and that all three can sit at the same table. Sections 9 and 20 test that claim directly; everything between them explains why it can be true at all.

2 · The one design choice

In every other memory game, the board gets easier as you win. Cards are removed. The unknown shrinks. A player who does well early is carried by the game itself.

In Limbo Match the matched red chips **stay face-down on the dial**. They look exactly like the ones still in play. Nothing is ever returned to the player's eye.

This single choice inverts the whole curve.

- The **visible** difficulty stays flat from the first pair to the last.
- The **real** difficulty rises, because the set of live reds shrinks only inside the player's head.
- Nothing is given back. There is no coasting phase.

Two quantities are needed to see this, and the difference between them is the whole point. One is **what the board still hides**. The other is **what is still genuinely in play**. In Concentration these are the same number, so a single line draws both — the board does the player's bookkeeping. In Limbo Match they come apart from the first match onward: the gold line never moves while the teal line falls, and the shaded gap between them is what the player is holding that the board refuses to show. That gap is the difficulty.

That is why a medium category does not run out of room. The board never helps you.

There is a second consequence, and it is the one that makes the scoring honest. Because a spent red is invisible, **remembering which reds are spent is itself a skill** — a provable one. It is not free information handed over by the interface. A player who tracks it plays a different game from one who does not, on an identical-looking board.

That claim can be measured, and it was, for this document. Take a player with no memory at all and run them on the shipped rule: they score 15.7 out of 48 on an eight-pair board. Now change one thing — let matched reds be lifted off the dial, as every other memory game does — and the same player scores 25.2. **One rule is worth a sixty per cent swing in what the board hands over for free.** That is the size of the design decision, in points.

There is a third consequence, quieter but worth naming. Because the board never reports progress, it cannot be read over a player's shoulder. In Concentration a bystander can see at a glance how far along a game is and which cards are gone; here the state of play lives entirely in one head. The game therefore has no spectator shortcut and no accidental coaching — the only way to know where a game stands is to have been paying attention from the first flip. That is an unusual property for a game to have, and it is what makes a recorded replay worth watching rather than merely worth checking.

3 · The tiers of skill

The game is layered. Every layer sits on the same board, and a player can be at any layer at any age. This is the sense in which the title of this document is meant: the tiers are not only the three products, they are the levels a mind can occupy.

Tier	What the player does	What it costs the brain	What it pays
1 · Sight	Flips and hopes	Nothing	Chance only
2 · Recall	Remembers reds already turned	Working memory, grows with P	Large — the first real jump
3 · Elimination	Tracks which reds are **spent**	Constant load, high discipline	Certainty in the endgame
4 · Tracing	Follows the chain instead of guessing	Almost nothing — one address at a time	Consistency; clean sweeps
5 · Economy	Prices each flip, each limbo, each clover	Judgement under a clock	Bonus points

Tier	What the player does	What it costs the brain	What it pays
6 · Harvest	Forces limbos, then **remembers the queue they return in**	Two structures at once — the dial *and* the order	The ceiling itself rises
7 · Tempo	Does all of the above fast	Dual-task load	The tiebreaker

A child plays tier 1 and 2. An expert plays tier 6 and 7. The score separates them cleanly, on the same board, with no handicap system and no matchmaking.

Tier 6 is the one that surprised this analysis, and it was found by a game rather than by a model — see section 11.

Chess has exactly this property. One 8×8 board serves a beginner and a world champion. It is rare, and it is not an accident in either game.

4 · The flip is a currency

The scoring rule is short: a match pays **1 point, plus 1 point for every flip not used**.

Points = 1 + (F – flips).

So every flip costs exactly one point. The exchange rate never changes. This is what converts memory into economics: the player is not asked "can you remember?" but "how cheaply can you know?"

For the 8-pair World Record board (P = 8, F = 5, jackpot 48):

How the pair was found	Flips used	Points
Blind tap on the first action (Clover)	1	6
Found on the first flip	1	5
Found on the second flip	2	4
Found on the third flip	3	3
Found on the fourth flip	4	2
Found on the fifth flip	5	1
Not found	5	0

Two things follow. First, the difference between a good player and a great one is measured in single flips — the scale is fine enough to register small improvements, which is exactly what keeps a player climbing. Second, the top row is worth studying: the blind tap pays **one more** than the sighted flip. That is not a quirk. It is the game paying a premium for certainty, and section 10 shows how an expert collects it.

5 · The hidden mathematics: the game is a permutation

The reds sit on dial positions 1 to P, and each carries a value 1 to P. That is a **permutation**: a mapping from position to value.

Every permutation breaks into closed loops. Position 3 holds red 7; position 7 holds red 2; position 2 holds red 3 — and the loop closes. The dial is a set of such loops, invisible until you start turning chips.

This gives the game its key structural fact.

If you are hunting blue b, start at dial position b and follow the chain. Flip position b, see its value, go to that position, flip, and so on. You will land on the red carrying b at the end of the loop that contains b — and the number of flips you need is exactly the **length of that loop**.

So the question "will I find this pair within F flips?" is exactly the question "is the loop containing it shorter than F?"

Every player who has ever traced a chain has been doing cycle-following on a random permutation without being told so.

A worked example makes it concrete. Eight reds are dealt as below, and the player is hunting blue 3. The chain starts at dial position 3 and closes after four steps — so this pair costs four flips, and nothing the player does can make it cost fewer.

Step	Flip dial position	Red found there	Is it blue 3?	Next address
1	3	7	no	7
2	7	2	no	2
3	2	5	no	5
4	5	3	YES — found	—

Two reds in this deal sit on their own numbers — loops of length one, found on the first flip for the full bonus. The rest of the dial is one long chain. That mixture, different on every shuffle, is what makes two games of the same category feel nothing alike.

6 · The surprise: tracing does not raise your average — it changes your tail

Here is the result that is genuinely counter-intuitive, and it was verified two ways.

For a single pair, tracing the chain and flipping at random have **exactly the same chance of success**. The reason is a clean piece of mathematics: in a random permutation, the length of the loop containing a given element is uniformly distributed over 1…P — and so is the position of the target under a random scan. Same distribution, same expected score.

The simulation agrees. At P = 12 the two strategies match to within noise at every flip limit from 3 to 8.

So why trace at all?

Because the average is not what a record is made of. A record needs a **clean sweep** — every pair found, no miss, no limbo. Under tracing, all the pairs in short loops succeed together; the successes are locked to each other. Under random flipping they are independent, and independence is fatal when you need eight of them at once.

Board	Flip limit F	Clean sweep, tracing	Clean sweep, random	Ratio
6P	4	63.9%	8.9%	×7
8P	5	57.2%	2.3%	×24
10P	6	51.7%	0.62%	×84
12P	8	61.5%	0.82%	×75
16P	10	54.5%	0.06%	×908
21P	13	53.5%	0.005%	×10,698

At the top of the ladder the tracer is more than **ten thousand times** more likely to sweep the board — while scoring the same on an average day.

This is the deep structure of Limbo Match, and it has a famous relative: the *100 prisoners problem*, where following a chain lifts a hopeless chance to about 31%. Grego's game contains that problem inside it, as a playable move.

It also says something about what skill *is* in this game. Skill here does not raise your typical score by much. It raises the ceiling of what is possible on a good board, and it makes perfection reachable at all. That is precisely the shape of skill in chess, in music, and in sport — the average moves slowly, the summit moves enormously.

This result is easy to disbelieve, so it is worth saying how anyone can check it without a computer. Take a pack of numbered cards, shuffle it, and lay it out. Pick any card position and follow the chain: the number you turn tells you where to look next. Count the steps until the chain closes. Do it for every starting point and you will find the whole pack falls into a few closed loops, and that the number of steps a start needs is simply the size of the loop it belongs to. Now ask the two questions separately. *Is one particular card found within F steps?* — that is a coin weighted by loop size, and random guessing weighs the same. *Are they all found within F steps?* — that needs only the longest loop to be short enough, which is one condition instead of many independent ones. The whole result is in that difference, and the cards will show it in ten minutes.

There is a practical lesson in it for anyone trying to improve. Practising *memory* raises your average and will make ordinary games less scrappy. Practising *tracing* does something different: it raises the chance that a good board turns

into a perfect one. A player who wants records should train the second, and should expect their ordinary scores to move very little while their best scores move a great deal.

It also explains something a player might otherwise take as a plateau. Weeks of practice can leave the average almost where it was while the ceiling of what you manage on a good night climbs steadily — and since most sessions are ordinary, most sessions will feel like no progress at all. The progress is real; it is simply concentrated in the games that were already going well. Anyone tracking their improvement should watch their best result over a month, not their typical one over a week.

7 · Is the World Record ladder calibrated? Yes — to a coin flip

The WR ladder fixes one flip limit per board size: MF[P], Grego's own integer sequence, registered as OEIS A389262.

Two independent checks were run against it.

Check one — the limit tracks the longest loop. For a random permutation of P elements, the expected length of the *longest* loop can be computed exactly. Set that beside MF[P]:

P	MF[P]	Expected longest loop	MF ÷ P
6	4	4.04	0.667
8	5	5.30	0.625
10	6	6.55	0.600
13	8	8.42	0.615
16	10	10.30	0.625
21	13	13.42	0.619

MF[P] sits within one flip of the expected longest loop at every rung of the ladder, and the ratio MF ÷ P hovers around 0.62. That number has a name in mathematics: the **Golomb–Dickman constant**, 0.6243, the limit of the expected longest cycle as a fraction of n.

It is worth being clear about why that ratio is the interesting one. A flip limit is a promise: it tells the player how much room they have before the board is allowed to beat them. Set the limit at the *average* loop length and most boards become unplayable, because it is the longest loop that decides the game, not the typical one. Set it at the longest loop plus a margin and every board falls, so the category stops asking anything. The only interesting place to stand is exactly on the expected longest loop — where the board is beatable, but only just, and only by a player who wastes nothing. That is where this ladder stands, at every rung from three pairs to twenty-one.

Check two — what that means at the table. The probability that a perfect tracer clears the whole board with no miss at all, computed exactly:

P	5	8	10	13	16	18	20	21
Clean sweep	55.0%	56.5%	52.1%	53.8%	54.8%	52.5%	50.5%	53.5%

From 5 pairs to 21 pairs, the flip limit is set so that **perfect play succeeds about half the time**. The ladder is a coin flip at every rung.

That is the definition of a well-set difficulty ladder, and it is not easy to hit by hand across twenty categories. A limit much higher and skill stops mattering; much lower and the game becomes a lottery. The WR ladder sits on the knife edge, all the way up.

Two things follow that are worth separating. The first is a statement about the game: a player at the top of any World Record rung is, in a precise sense, in a fair fight with the shuffle. Neither side is favoured. Half the boards can be beaten outright by faultless play and half cannot, so a record attempt is never hopeless and never routine, and the difference between a good session and a great one is genuinely decided at the table.

The second is a statement about the ladder as an artefact. Holding one probability near a half across nineteen rungs, with an integer flip limit as the only control, is not something that falls out of a simple formula — the expected longest loop is not linear in P, and the rounding has to land the right way at every step. Whatever route produced this

sequence, the result behaves like a designed instrument rather than a rule of thumb, and it is the reason the twenty categories feel like one coherent ladder instead of twenty unrelated puzzles.

8 · The measured ladder of players

To put numbers on the spectrum, five players were simulated on the 8-pair World Record board (P = 8, F = 5, L = 1, jackpot 48). Each one adds exactly one ability to the one before. 40,000 games each, under the shipped rules.

Player	Mean score	% of jackpot	Spread (sd)	Best seen
A · Blind luck, no memory	15.7	33%	5.1	37
B · Tracing, still no memory	15.0	31%	10.0	40
C · Recall, random order	36.6	76%	1.9	40
D · Recall and tracing	36.6	76%	1.9	40
E · Plus elimination and endgame clover	37.5	78%	2.0	41
— · Jackpot ceiling	48	100%	—	48

Before reading the numbers, note what the five players are and are not. They are not attempts to imitate real people; they are five clean hypotheses, each isolating one ability so its worth can be priced on its own. Nobody plays like player A, and nobody has player C's flawless recall. The value of the ladder is in the *steps between the rows* — what each single ability adds when everything else is held still, which is exactly the quantity a real player cannot measure for themselves.

Read this table carefully, because it contains four separate lessons.

Memory is the great divide. From 15.7 to 36.6 is the single largest jump available to any player. Everything else in the game is decoration next to it.

Tracing does not show up in the average — exactly as section 6 predicts. But look at the spread: player B's standard deviation is *double* everyone else's, and B's best game beats A's best by three points despite a lower mean. Tracing buys the tail, not the middle.

Elimination and the endgame clover are worth about one point. Small — but records are won by one point, and this is a skill that a determined player can execute every single game.

The gap to the ceiling never closes. The best simulated expert scores 78% of the jackpot. The remaining 22% is not reachable by skill; it needs blind luck as well (section 21). The summit stays open forever.

It is worth being clear about what these five players leave out, because the omissions all run the same way. None of them uses the Pin. None of them takes the zero-cost gamble at two blues. None of them plays the first blue as reconnaissance. Each of those is a real technique available in the shipped game, and each would raise the top of this table rather than the middle of it. The ladder above is therefore a floor on what skill is worth, not a ceiling — which is the safe direction for a model to be wrong in.

One number in that table is worth dwelling on. Player C and player D score the same to two decimal places, and they are not the same player: one searches at random, the other follows the chain. Identical means, different methods — the section-6 result appearing again, this time from the other end. If a reader takes one thing from the simulation it should be this: in a game that looks entirely like a test of memory, the order in which you look turns out not to matter to your average at all. What matters is whether you remember, and then whether you can convert a good board into a perfect one.

9 · Why every age can climb: three ways to hold a board

Human working memory holds roughly four items at once. This is among the most robust findings in cognitive psychology, and it does not change much with age or training. What changes is not the number of slots but what a player can fit into one. That is the classic explanation for why chess masters recall a real position at a glance and a random one no better than a novice: the master is not holding more pieces, but fewer, larger units.

Limbo Match can be held in a mind by three different routes. They are separate faculties, they can run at the same time, and — the point of this section — they do not decline together.

Route	What is actually stored	Where it runs out	Who it favours
Auditory — rehearse the sequence	An inner voice repeating "three, seven, two"	About four items, and it needs quiet	A rested mind, undisturbed
Visual — the look of the board	Where a chip sits on the dial, the shape of an icon or numeral	Its own limit, but a separate one — it does not compete for the loop	Strong visual imagers; the KIDS tier leans on it
Semantic — chains of meaning	1-10-9-4 as one idea rather than four numbers	Effectively unbounded — it is limited by what you know	Whoever has lived longer and read more

The third route is the one that breaks the ceiling

The first two work *inside* the four-item limit. The third changes what an item is. A player who sees 1-10-9-4 and thinks *God, the Commandments, nine months, the four letters of the Name* has spent one slot, not four. A football supporter reading 2-11-12-3 as *two halves, eleven players, a clocked match, three officials* has done the same. 5-7-8-6 is a year in the Hebrew calendar. Thirteen sits in the middle as the number everyone already has a feeling about.

This is **chunking**, and it is the whole reason the top of the ladder is playable at all. Nobody rehearses twenty-one numbers. Five meaningful groups is within ordinary reach. Without chunking, 21 pairs would be a wall; with it, it is a long climb.

And chunking is the one route that **improves with age instead of despite it**. Rehearsal capacity narrows over a lifetime. The store of associations only grows. A sixty-year-old has more hooks to hang a number on than a twenty-year-old, and 5786 means nothing at all unless you have lived somewhere it mattered. That is an unusual property for a memory game. Most of them reward the young; this one rewards whoever has more furniture in their head, and lets them convert it into points.

Why tracing and chunking help each other

The permutation structure of section 5 suits this unusually well. A traced loop *is* a sequence, in a fixed order, arriving already shaped as a chunk — so the technique that costs the least memory is also the one that produces the most memorable material. Tracing needs one address at a time whatever the board size, and what it hands back is exactly the kind of ordered chain a semantic memory likes to keep.

Route	What must be held	How it scales with board size
Remembering every red	Up to P position–value pairs	Badly — beyond most people by 12 pairs
Following the chain	One address at a time	Not at all — flat in P
Chunked chains	A few meaningful groups	Well — the units grow with the player

A traced chain that passes through reds you already know costs **no flips at all** — remembered ground is crossed for free. So memory does not merely add to tracing; it multiplies it, and a chunk multiplies it again. That compounding is what a player feels as endless depth.

What this means at three different ages

The game does not have one entry requirement. It has several doors into the same room, and which one a player uses depends on what their mind is currently good at rather than on how old they are.

Player	What they bring	What the game gives back
A young child	A small span, quick pattern learning, no reading yet	Iconic chips are numbers in picture form, so a pre-reader plays the real game — not a simplified one
A teenager or adult	Full rehearsal capacity, appetite for optimisation	The whole ladder: loops, flip economy, the clover window, the pin
An older player	A narrower loop, and a lifetime of associations and strategy	Chain-following, which needs one number at a time, and chunking, which turns what they know into capacity

That third row is the one most easily missed. In an ordinary memory game a narrowing span is simply a disadvantage, and nothing in the design offers a way round it. Here there are two ways round it, and neither is a concession or a handicap setting — they are the strongest techniques in the game. An older player who traces loops and chunks them is playing better than a younger one who does neither, and the score says so without anybody having to be told.

What the game therefore asks for, and what it does not

Two practical consequences follow, and both match what strong players report. **It asks for quiet.** The auditory route is disrupted by background speech and music even when they are ignored — a well-replicated effect — so the same player will score worse in a noisy room through no failure of skill. **It asks for patience.** There is no reward for hurrying: time is only a tiebreaker, so a rushed search does not buy anything, it just spends flips and loses points outright.

It is worth stating what the game does *not* ask for, because it is easy to assume the opposite. It does not ask a player to be fast. Speed here is not a strategy that trades against accuracy; it is a by-product of having the board properly encoded. A player who knows where a red is does not hesitate, and one who does not know cannot hurry usefully. The quick player and the accurate player are, almost always, the same player — which is why the ranking treats time as a tiebreaker and never as a currency.

One more property belongs here, because it widens the door further than age alone. The game asks for no reading and no language. A chip is a number or a picture; the dial is a ring; the score is a count. Nothing has to be translated, and nothing depends on vocabulary, on schooling, or on being a fluent reader of any particular alphabet. A grandmother and a grandchild who share no common written language can still play the same board and compare the same numbers. The semantic route is personal — each player brings their own meanings — so it needs no shared culture either, only a lived one.

A caution belongs with all of this. The three routes are well-established features of human memory, and the game plainly engages them; that much is uncontroversial. Which route any individual player actually leans on has not been measured here, and could only be settled by watching real people play. What follows is offered as mechanism, not as a finding.

10 · The Clover: a lottery for the novice, a payout window for the expert

The Clover looks like the luck button. It is the opposite, and the mathematics is exact.

Pressing Clover means betting blind on a hidden blue. If **k** blues remain and you know where every live red sits, your chance of hitting is 1 ÷ k. Work out both branches and the expected value is:

Clover = F – **1 + 2 ÷ k**, against **F** for simply flipping the red you already know.

The difference is **2 ÷ k** – **1**:

Blues left (k)	Clover advantage	Verdict
1	+1 point	Always clover. Free point.
2	0	A true coin flip, at no cost.
3	–0.33	Do not.
4 or more	–0.5 or worse	Do not.

Three consequences, each of them a genuine strategic rule.

The last blue must always be clovered. A player who knows the board and flips it instead has thrown away a point. That is a rule a beginner will never guess and an expert will never miss.

At two blues the game offers a free gamble. Zero cost, pure variance. When a record needs one more point, you take it; when you are already clear, it does not matter. A game that hands the player a genuinely balanced choice, with no right answer, is a game with depth.

This is why the fan matters. The fanned chip under the top blue is the signal that the two-chip window has arrived. It is not decoration; it is the interface telling the player that the payout table above is now live.

And note what the Clover does to the whole design: the same button is a lottery for a player who knows nothing and a certainty for one who knows everything. Luck and skill are not separate systems in this game — they are the same lever, read differently by different minds.

Compressed into something a player can carry to the table: never flip a red you can clover for the last blue; treat two blues as a free coin toss and take it when a record is one point away; and at three or more, simply play. Three sentences, exactly derived, and worth roughly a point a game — which over a season of record attempts is a great deal more than it sounds.

11 · The Limbo: failure converted into information

A missed blue can be sent to the back of the stack to return once, at the tail. Three details make this more than a mercy rule.

It is positional. Only the first L blues are eligible. In the World Record game L = 1, so exactly one blue in the whole game can be forgiven — the first. But L is a category setting, not a constant: custom categories run to L = P, and everything below changes character as L rises.

It is one-shot. A returning blue gets no second reprieve.

It requires virginity. No limbo is granted if any red was turned twice during the first search. This is the clause that stops brute force: a player who flails and re-flips forfeits the safety net. The limbo is reserved for a clean, disciplined search that simply ran out of room.

Together these turn a miss into a *decision*. Because the first blue is the only forgiven one, an advanced player can treat it as reconnaissance: spend the flips learning reds rather than chasing the match, accept the limbo, and cash the blue at the tail when its red is already known. The cost is the first blue's bonus. The gain is a map.

That is a real opening theory, in a memory game, arising from a single rule. And the virgin rule is what keeps it honest — the scouting run must still be clean.

The harvest: the strategy that raises the ceiling

Each limbo Cyrus can spend is worth exactly one more point on the ceiling, until the board runs out of blues worth banking. Computed on the deal of a real record (P = 8, F = 5).

There is a second use of the limbo, and it is the deepest single idea in the game. **It was already documented** — `limbomaths.pdf` names it *the deliberate-limbo harvest* and states the mechanism exactly: a spare slot forces a known blue down to be clovered on return for 1 + F instead of F, *"at the cost of carrying its value and position across the wait"*. What is new here is not the discovery but the omission: **the first edition of this document missed it**

entirely, and a real game is what sent me back to look.

A returning limbo blue can be taken with the **Clover** instead of a flip. Compare the two ways to finish it:

	Points	What it requires
Flip to the red you remember	1 + (F – 1)	The red's position
Clover the returning blue	1 + F	The red's position **and** the blue's place in the return queue

One point more, per blue, for one extra thing remembered. And the Clover is not a gamble here: the blue's identity is known because you watched it go to limbo, and the red's position is known because you flipped it earlier. **Nothing is being risked.** It looks like the lottery move and it is in fact the purest memory move in the game.

So at higher L a player can deliberately *fail* blues they already know, banking them for a richer payout on return — spending points now to collect more later. On the eight-pair board of a real record, that line is worth **+5 on the ceiling**: 36 without the harvest, 41 with it. The ceiling model already knows this and plays the line; the code comment on the branch that does it reads simply *"force it down: 0 now, 1+F later"*.

What makes it a genuinely separate skill — the sixth tier of section 3 — is that it demands **two structures at once**. The dial is a map, and it is stable. The limbo queue is a sequence, and it *forms during play*, growing as the game goes wrong.

One honest qualification, and it comes from the same source. The queue is not held on memory alone: limboed blues return **in order, yellow-ringed and fanned**, so the interface shows what is coming. The demand is therefore lighter than pure recall — but it is not free either, because a player still has to have watched the queue form and to hold it *against* the dial while playing. Reading the fan and having tracked its contents are different things, which is why a player can be excellent at reds and still leave the harvest on the table.

That is not hypothetical. It is exactly what happened in the game that prompted this section: every red remembered, all six limbos forced, and not one of the six return positions held. Six points left on the board — and, as section 17 explains, a title with them.

Pedagogically the limbo does something else. It brings a failed item back after a delay, once, at the end. That is spaced repetition, inside a single game.

For a young player that is worth more than the point it saves. A miss in most games is simply a loss, and the lesson it teaches is that missing is bad. Here a miss comes back once, later, when more of the board is known — so the lesson it teaches is that a failed attempt still left information behind, and that the second look is easier because of the first. That is a true statement about learning anything, delivered by a rule rather than by a message on a screen.

12 · The Pin: the only move that changes the world

Every other action reads the board. The Pin **rewrites** it. Tap a red and the whole ring rotates until that red sits on its own number. The chips keep their values and their order; only their positions shift.

Mathematically this composes the permutation with a rotation — which produces an entirely new loop structure while destroying none of the player's knowledge. A long loop can become several short ones. A dead board can become a winnable one. And it costs **no flips at all**.

How strong is it? For each board there are P possible rotations. Simulation of 20,000 boards per size:

Board	F	Winnable as dealt	At least one rotation is winnable	Share of rotations that work
8P	5	56.3%	~100%	56.5%
10P	6	52.0%	99.9%	52.0%
16P	10	54.8%	~100%	54.8%
21P	13	53.5%	~100%	53.5%

Read the third column again. **A board that cannot be swept by tracing can almost always be rescued by a pin** — if the player knows enough to choose the right one. Around half of all rotations work, so the chance that none of twenty-one works is vanishingly small.

There is a memory dividend here as well, which follows from section 9. Because a pin keeps the reds in the same cyclic order and only moves their positions, any chunk a player has already built survives it — the chain stays the same chain, sitting on different dial numbers. A tool that reshuffled would destroy the player's work; this one moves the board and leaves the memory of it intact.

A pin therefore does something no other action can: it changes the difficulty of the board that remains, without spending a flip and without costing the player any of what they already know. Every other tool trades one resource for another. This one asks only that you understand the board well enough to choose.

This is the deepest lever in the game, and it is beautifully self-balancing: the Pin's power is exactly proportional to how much of the board you know. A blind pin is a free re-roll — worth taking on a dead board, worth nothing to someone who cannot tell a dead board from a live one. A sighted pin, in the hands of a player who has traced the loops, is close to a guarantee.

If the Clover is this game's endgame technique, the Pin is its castling: rare, structural, and decisive.

13 · The Eye: the game prices its own cheat at infinity

An Eye under the blue stack shows every red. Using it voids the submission.

Most games hide the possibility of cheating. This one puts it on the board, in the open, and sets its price at everything. The result is a tool that is genuinely useful for learning and practice, and worthless for records — so it can be offered freely without damaging the ranking.

It is also a small lesson in integrity, delivered without a lecture. The player is trusted with the key and shown exactly what using it costs.

14 · Time: the second axis, and why the spectrum feels endless

Score measures economy. Time measures fluency. In the ranking, time separates equal scores.

This matters more than it first appears. A player who has reached their score ceiling in a category has *not* reached the end of the category — they can still get faster, and getting faster while holding accuracy is a different and harder skill. It is a dual task: recall and deduction under a clock.

That is the honest answer to Grego's "almost infinite". A single axis has an end. Two axes, where the second is fluency, do not — because fluency improves continuously and never quite finishes. Musicians know this exactly: the notes are learned long before the piece is.

One thing this is **not** is a trade between speed and accuracy. There is no dial a player can turn toward haste and away from care: hurrying does not buy points, it spends flips. Time falls as a consequence of knowing the board, not as an alternative to knowing it. That is why the ranking puts score first and lets time separate equals — the ordering matches how the two quantities are actually produced.

	Score	Time
What it measures	Economy — how cheaply you knew	Fluency — how readily you knew
How it is earned	Unused flips, banked as points	Hesitation removed
Ceiling	Hard: the jackpot, P × (F+1)	None in practice
Rank weight	Primary	Separates equal scores
Improves by	Understanding	Repetition

The ordering matters as much as the two axes themselves. Score is primary and time only breaks ties, which keeps the game about knowing rather than about hurrying. A player cannot buy rank with speed alone — but between two players who both played perfectly, the one who hesitated less is ahead. That is the correct order for a game whose subject is concentration.

15 · Worth: what a category is really worth

Every category in the game carries a number between 1 and 441. It is called its **worth**, and it is not the jackpot, not the board size and not the flip limit. Worth answers one question, and it is the right question: **how much of this category's score has to be earned?**

The answer needs two measurements of a shuffle, and they are the heart of the whole system.

	What it is	Who could take it
Floor	What this exact shuffle hands over free	A player with no memory at all
Ceiling	The most that can be taken without ever gambling	A player with perfect memory
Room	Ceiling minus floor	What is actually at stake

Worth is the largest room any deal inside the category can offer — the maximum of (ceiling − floor) over every loop shape the category's window allows, with a minimum of one.

Why worth is not the jackpot — the example that killed the old measure

Category	Jackpot	Floor	Ceiling	Room = worth
21 pairs · 13 flips · Unit length	294	273	274	1
13 pairs · 8 flips · Longest loop	117	0	103	103

The first line looks like the hardest thing in the game: twenty-one pairs, a jackpot of 294. But *Unit length* means every red sits on its own number, so every loop has length one and the first flip finds every pair. The shuffle gives away 273 of the 274 available points to a player who remembers nothing whatever. There is nothing to prove there. Worth 1.

The second is a smaller board with a lower jackpot — but its reds form one long chain, so a memoryless player scores nothing at all. Every point has to be earned. Worth 103.

The retired difficulty measure D = P ÷ (F + L + 1) rated these two within about a tenth of each other. Two categories that are worlds apart, scored as twins. That is why D was retired from merit ranking, and why worth replaced it.

Worth is a lower bound, not an exact figure — a correction

An earlier version of this document claimed worth was exactly computable, on the grounds that a deal's ceiling depends only on its loop shape and never on the order the blues arrive in. **That claim was tested again for this revision and it is false.** The correction is set out here rather than quietly removed, because the way it failed is more interesting than the claim it replaces.

The original test varied the blue order while holding the deal fixed, and found zero spread — correctly. What it never varied was **the deal within a shape**. Doing so at five pairs, exhaustively, over all 120 arrangements: **five of the seven loop shapes produce two different ceilings.** Shape `5` gives 16 or 17. Shape `1,4` gives 16 or 17. The shape does not determine the answer.

A second and larger effect appears once the harvest of section 11 is available. With no limbo to spend, blue order genuinely does not matter — zero spread, as originally found. With limbos to spend it matters a great deal: on one real record's deal, 300 blue orders produced **five different ceilings, from 37 to 41**, because which blues can profitably be banked depends on the order they arrive in.

Worth is computed by sweeping loop shapes and testing one representative deal for each. Both effects are therefore invisible to it, and both push the same way. Measured against a search over many deals and many blue orders:

Category	L	Worth as computed	Found by search	At least
5·3·1 (WR rung)	1	16	17	+1
8·5·1 (WR rung)	1	41	41	0
13·8·1 (WR rung)	1	103	103	0
10·6·4	4	61	63	+2
8·5·6	6	23	27	+4

Category	L	Worth as computed	Found by search	At least
9·6·7	7	55	61	+6

The pattern is orderly: **the World Record ladder is essentially sound** — ten of eleven rungs tested match exactly, because at L = 1 there is at most one harvest to find. The understatement grows with the number of limbos, which is precisely where the harvest becomes a strategy.

Three things follow, and they should be read together.

The error only ever understates. Worth is a maximum found by searching; missing candidates can only make it too small. No category is rated or priced above its true standing.

No player's merit is affected. GLMP is computed from the shuffle and the blue order the player actually faced, never from the category's worth. Every score, every rank and every record decision in the game is unaffected by this. Worth is used for sorting, for tournament difficulty percentiles and for sponsorship prestige — descriptive and commercial work, not competitive work.

Exactness is not recoverable. Now that the shape is known not to determine the ceiling, an exact worth would require enumerating every deal and every blue order: 6.2 billion × 6.2 billion at thirteen pairs, and 5 × 10¹■ of each at twenty-one. Worth is therefore, and permanently, **a conservative lower bound on what a category can hold**. A better search will raise some values; nothing will make them exact.

This is the discovery that this revision is proudest of, for an uncomfortable reason: it was not found by checking the model. It was found because a live record — LimboFan's 44 of 48 in an eight-pair, six-limbo category — **achieved a merit of 27 in a category whose theoretical maximum was computed as 23.** A player beat the ceiling the system had claimed for the category, and the ceiling was wrong. That is the sort of thing a document like this exists to catch, and it took a real game to catch it.

The game computes worth server-side at submit time, because a submit is rare and already doing work, while a Champs render is not.

Worth then does real work elsewhere in the system. It sorts the houses in the Champs table, so a category's place is set by what it demands rather than by how big its numbers look. It prices tournaments: the organiser's difficulty slider is a percentile that maps straight onto worth, so a tournament advertised as "Competitive" means something exact. And it sets sponsorship prestige, so a sponsor pays for the standing of a category rather than for its headline size. One measurement, three uses, all of them honest for the same reason.

It is worth noticing what worth is *not* used for, since the omission is deliberate. It does not scale anybody's score, it does not weight a leaderboard, and it never multiplies or divides a result. A category's worth says how much merit that category can hold; the merit a player actually earns is computed from their own shuffle, by subtraction, with no reference to the category at all. Keeping those two apart is what stops the system from double-counting difficulty — the mistake that sinks most attempts to compare results across categories, and the one the retired D measure was heading toward.

Across the whole Feast of 220,407 categories, worth runs from 1 to 441 with a median of 92: a quarter of all categories are worth 46 or less, and only one in a hundred is worth more than 337. The nineteen World Record rungs run 8, 14, 16, 25, 36, 41, 55, 60, 77, 96, 103, 125, 133, 158, 185, 195, 225, 236, 269 — climbing steadily, but nothing like as fast as the jackpot beside them.

16 · Floor and ceiling: measuring one game

The same two measurements, applied to the shuffle a player actually faced, give that game's merit:

$$GLMP = \max(1, \min(S, \text{ceiling}) - \text{floor})$$

Read it from the inside out. **Subtract the floor** — a player gets no credit for what the shuffle handed over. **Cap at the ceiling** — anything above what perfect memory could have taken was luck, a blind clover no amount of knowing could have supplied, and it is capped away. The ceiling is a cap, never a divisor. **Lift a zero to one** — Grego's rule: finding the game, learning it and scoring high enough to hold a record anywhere already takes some brains, so nobody who holds a record scores nothing.

The blind run — one case the formula above gets wrong

That formula is right for a player who reads the board, and wrong for one who never looks at it. Before the first flip no red has been revealed, so a leading run of clover hits is **provably** uninformed — the log shows it and nothing has to be inferred. This is not the undecidable question of whether a particular clover was informed; it is a count of flips.

For such a player the board is irrelevant, and it genuinely does not help them: an all-blind run comes off at 1 in P! whatever the shuffle. Measured over 200,000 boards at six pairs, the rate is 1 in ~720 in every floor bucket, from floor 0 to floor 22. Charging them the trail-follower's floor deducts points that were never free to them, and the memory ceiling measures a route they never took. On the thirteen-pair rung that made one identical feat worth anywhere between 6 and 103 — a seventeen-fold swing on a property with no bearing on the result.

The only skill in a blind run is **elimination**. Matched reds stay face-down and look exactly like unmatched ones, so remembering which are spent is what turns a 1/P guess into 1/(P – j): worth 26 times at five pairs and 114 million times at twenty-one. So the run is measured against blind bounds —

$$\textit{blind floor} = (1 + F) \cdot n/P \cdot \textit{blind ceiling} = (1 + F) \cdot \Sigma\, 1/(P - j),\ j < n$$

— and the usual clamp does the rest, since scoring above what perfect elimination expects is luck. From the first flip the game continues as a genuine (P – n)-pair board, scored the normal way.

A blind jackpot is therefore worth 2 merit at three pairs, 4 at five, 20 at thirteen and 37 at twenty-one. Small, scaling with the board, and far below the hundred-odd that reading a shuffle earns. **A lucky player may hold a high seat and still rank low.** That is the answer to a real worry: that it is pointless to think hard against people who are guessing.

How the floor is computed

The floor is the score of a player with no memory who simply follows the trail. For a red sitting in a loop of length c, that walk takes exactly c flips — and it takes c flips for *every one* of the c pairs in that loop. So each loop contributes c × (1 + F – c) points, and loops longer than the flip limit contribute nothing at all:

$$\textit{floor} = \Sigma \textit{ over loops with } c \le F \textit{ of } c \times (1 + F - c)$$

It depends only on the shuffle and the flip limit — not on the order of play, not on the player. It is precisely the gift. A worked example, on an eight-pair board with five flips:

Loop length	Pairs in it	Flips each	Points each	Contributed
1	1	1	5	5
1	1	1	5	5
3	3	3	3	9
3	3	3	3	9
Floor				**28**

Twenty-eight of that board's forty-eight points are simply given away. A player who scores thirty on it has earned two. The same thirty on a board whose reds form one long chain would have earned all thirty.

How the ceiling is computed

One design decision inside the ceiling deserves to be stated, because it shows what the system is protecting. The ceiling deliberately errs **high**: where two perfect-memory lines are possible, it takes the better one. An overstated ceiling costs nothing — it simply caps less luck than it might have. An understated one would be a real injustice: it would brand genuinely great play as luck and cap away merit the player had earned. Faced with an asymmetric error, the model chooses the side that can only be generous.

The ceiling is a full line of perfect-memory play under Grego's rules, taking the best result. It knows all it has seen, deduces the last red once the others are accounted for, uses the Clover only when both blue and red are known, may spend a limbo slot to clover a known pair, and may use any Pin orientation. It never gambles.

The anti-farming property was measured rather than assumed. Across 60,000 shuffles, a fixed-skill player's raw score turned out to be independent of how kind the deal was (correlation +0.00) — with memory, the flips a game costs are

governed by information, not by loop shape. Their GLMP, however, moved almost exactly opposite to the floor (correlation –0.99). In plain terms: **the same performance on a generous board earns almost no merit, and on a bare board earns nearly all of it.** That is the intended behaviour, and it needs no eligibility rule to enforce it.

Three consequences fall straight out of the arithmetic. There is **no category term** — no D, no √P, no jackpot — so the model never has to decide how hard a category is; it only asks what that shuffle offered. **Farming is impossible by arithmetic rather than by rule**: a kind shuffle has a high floor, so grinding an easy board earns almost nothing, and no eligibility gate is needed to say so. And **luck is removed rather than punished** — a lucky blind hit still scores points on the board, it simply does not count as merit.

17 · From one game to one player: GLMR and GLM100

A player is ranked by their **best single GLMP**, never by a sum. Grego's phrase for the rule is "merit, not marketing": no amount of playing can substitute for one game played well. It is the same principle as a personal best in athletics, and it has the same effect — volume buys nothing.

Rule	What it means
Best single GLMP	One great game defines a player, not a thousand ordinary ones
GLMR = 1 + (nicks strictly above)	Competitive joint ranks: 1, 2, 2, 4 — no tiebreak, equal merit stays equal
Every measurable nick is ranked	Not only the top hundred
Coupling rule	A nick is credited only for categories it actually holds
Not measurable = not ranked	Old records without a stored shuffle get a dash, never a guessed rank

The measurement in section 16 gives this rule a sharper meaning than it first appears to have. Because merit swings with the deal, a single GLMP describes a game rather than a person — so ranking by the best one is not merely a nice principle, it is what turns a per-game number into a statement about a player. And the statement is precise: a high standing does not mean "plays well on average". It means **once met a demanding shuffle and took nearly everything it had to give**.

That last rule deserves a note, because it is the kind of discipline rating systems usually fail. Records logged before the engine stored the deal cannot be replayed, so no floor and no ceiling exist for them. They are shown unranked. A system that guessed a value there would be easier to use and quietly wrong; this one refuses.

Why the best single game, and not a total

Every alternative was available, and each fails in a way worth naming.

If a player were ranked by…	What it would actually reward
The sum of all their GLMP	Time spent. A patient grinder outranks a better player.
Their average GLMP	Caution. The best move becomes to stop playing after a good run.
Recent form	Availability. A player with a free month outranks one with a job.
Their best single GLMP	The best they have ever played. Nothing else.

The last line has one more property that matters for a game meant to last: a personal best can never be taken away by somebody else's play, and it can never decay. A player who set a great game two years ago still holds exactly the merit they earned. The board can be climbed past, but not eroded — which is the opposite of how Elo behaves, and much closer to how athletics treats a record.

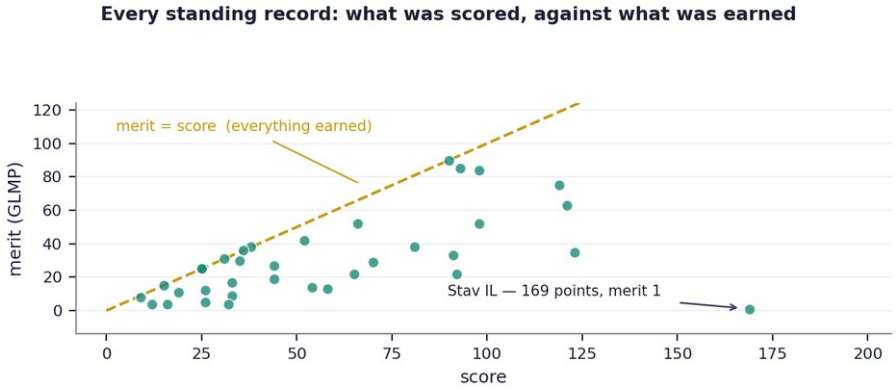
The coupling rule keeps that honest in the other direction. A player is credited only for categories they actually hold, so a great game whose record has since been beaten no longer carries the nick up the list. Merit is permanent as a fact about a game; standing is not. That is the correct pairing — the achievement cannot be taken away, but the position on the board still has to be defended.

Since 27 August that defence has two fronts rather than one: a seat can be lost to a bigger score, or to a better-played game that scored less. The pairing below sets out how the two kinds of champion are exposed to each other, and it is more balanced than it first looks.

Seat held by	Vulnerable to	Protected by
A lucky, high-scoring game	Merit — its GLMP is low and easily beaten	Its time, usually quick
A clean, well-played game	Time — a faster player can take it	Its merit, hard to beat

Neither kind of champion is safe, and neither is safe in the same way. That is what keeps a belt moving.

Merit can take a title — the three-way rule



Every measurable standing record. The dashed line is merit = score — everything earned. Each record falls below it by exactly what its deal gave away.

Everything above describes how merit is *measured*. From 27 August 2026 it also decides who holds a record, and the change is worth stating carefully because it is unusual.

A category has one seat. Until now it went to the highest score, with time breaking a tie. That produced a quiet injustice, and the numbers show its size: **in 26% of pairs of standing records, the higher score carries the lower merit.** The extreme case in the live store is a record of 169 points whose merit is 1 — a very large score that, measured against its own board, is almost entirely gift and luck.

Worse, the injustice compounded. Merit can only be computed from games that are *in* the store, and games only entered the store by beating a score. **A player's finest game was invisible unless it also happened to be their highest-scoring one.** The merit board's input was filtered through a score gate.

There are now three ways to take a seat:

	Condition
1	Higher score
2	Equal score, faster time
3	**Higher merit, with time no slower than the holder's**

The third is the new one, and two things make it sound rather than sentimental.

It cannot be gamed by scoring low. GLMP rises with score right up to the ceiling and only clamps above it. So merit and score agree until the ceiling and diverge only beyond it, where the surplus is gambling by definition. Checked against every measurable record in the live store: **merit never exceeds the score.** To beat a seat holding merit *m*, a challenger must score at least *m* plus their own floor. A weak score cannot carry a high merit, and the record cannot collapse to something trivial.

The time condition is load-bearing. Without it a player could grind a high-merit game slowly. Time is the one quantity luck cannot manufacture, and it was already the tiebreaker. The condition is *no slower* — faster or equal.

A natural worry is that this favours the slow thinker over the quick one, since careful play takes time while a lucky clover ends a blue instantly. The worry is reasonable and **this document cannot settle it.** The obvious test — do merit-heavy records run faster or slower per pair? — was run on the thirty-five measurable standing records and the answer was noise: the middle third came out slower than either extreme, board sizes differ systematically between the groups, and seconds-per-pair by board size ranges from 5.8 to 19.4 with two or three records apiece. Thirty-five games across nine board sizes cannot answer it, and reporting the two extremes while omitting the middle would have produced a tidy finding that the data does not contain.

What can be said is structural rather than empirical. The time condition does not need the correlation to be justified: it exists so that merit cannot be ground out slowly, and time was already the tiebreaker between equal scores. Whether quick play and merit tend to travel together is a genuine open question, and it belongs with the others in section 25 — it needs play data, not more analysis.

The consequence for the record book is the interesting part. **A record score can now go down.** A seat taken on merit may carry fewer points than the one it replaced. This is deliberate, and it is what the system is for: a seat is a *belt*, not an archive of the highest number ever reached. A belt that only ever ratchets upward eventually becomes unreachable — and since the jackpot cannot be taken without luck, a purely score-based ladder ends as a contest of volume. A belt that merit can take back does not have that ending.

Grego put it in a sentence that is hard to improve on: **the jackpot cannot be hit without luck, and merit can always take its title.** The nearest analogy is tennis, where a point won by an opponent's unforced error and a point won by a clean passing shot look identical on the scoreboard — except that here, only one of them counts toward the belt.

18 · Why this is not Elo — and may well be new

Elo and Glicko are the reference points, and Grego studied both before building this. They are excellent at what they do, and what they do is a different job.

	Elo	Glicko	Worth · GLMP · GLM100
What it measures	Strength relative to other players	The same, plus confidence in the number	How much of a specific shuffle the player earned
Needs an opponent	Yes	Yes	No
Needs a history	Yes — a connected graph of games	Yes, and it decays with inactivity	No — one game is enough
Affected by others	Yes — your rating moves when they play	Yes	No — a game's merit is fixed forever
Luck	Averaged out over many games	Averaged out, with a stated uncertainty	Subtracted and capped, per game
Provisional period	Yes	Yes, explicitly modelled	None
Scale	Open-ended, drifting	Open-ended, with a deviation	Fixed 1 – 441

Elo answers "how do you compare with other people?" It cannot answer "how well did you play *that* hand?" — and in a game where every board is a fresh random deal, the second question is the one that matters.

The nearest relatives, and what is missing from each

System	What it does	What it lacks here
Golf par	A fixed target per hole, same for everyone	Par is per hole, not per instance of the randomness
Duplicate bridge	Removes luck by having many pairs play the identical deal	Still relative: it needs a field playing the same board
Chess engine accuracy	Measures a player against the best line in the actual position	A ceiling only — a chess position gives nothing away free

Chess accuracy is the closest relative: it is instance-specific and it does compare a human with a perfect player. What it has no need of — and what Grego's system adds — is the **floor**. A chess position hands you nothing; a shuffled dial hands you a great deal, and it hands a different amount every time. Subtracting that gift is the genuinely unusual half of this design.

An honest verdict, with the hedge stated plainly. I cannot prove a negative about the published literature. But I know of no rating system that measures a player against *both* a computed floor and a computed ceiling of the specific random instance they faced, produces an absolute per-game merit on a fixed scale, needs no opponent, no field, no history and no provisional period — and then reuses the same two measurements to price every category in the game. The individual pieces exist. This combination, I have not seen.

Whether or not it is unprecedented, it is demonstrably **fit for purpose in a way Elo is not**, because it answers the question this game actually asks.

There is a second reason the choice suits this game in particular, and it is about who plays it. A relative rating needs a pool: enough players, playing each other often enough, for the numbers to mean anything — and a pool of mixed ages and abilities is exactly the case where a relative rating behaves worst, because a beginner's first ratings are noise and a child who plays twice a month never stabilises. An absolute measure has no such problem. A nine-year-old who plays one beautiful game on six pairs gets a real number for it, immediately, computed from that game alone, and it stands beside an adult's number without either needing a separate table. For a game meant to be played across generations, that is not a technical nicety; it is the difference between one community and several.

It also removes the quiet unfairness that relative ratings carry for anyone who plays irregularly. In an Elo-style pool, a player who disappears for a season returns to a number that has been moved by other people's results, and a player who plays constantly accumulates an advantage that is partly just presence. Here a record set on a Tuesday in March is worth exactly the same in December, and worth exactly as much to somebody who plays once a month as to somebody who plays every day. For a household where one person is enthusiastic and another dips in occasionally, that is the difference between a shared game and a private one.

19 · The dial, and the opponent who is always there

A category is a shape: pairs P, flip limit F, limbo allowance L, and the loop window. Together these define 220,407 distinct categories — nineteen World Record rungs and 220,388 custom ones, each with its own worth.

Flow — the state where a task is neither boring nor frightening — requires the challenge to sit just above current skill. Chess reaches it through an opponent of matching strength, which means chess needs a matchmaking system, a rating pool, and somebody else who is free right now. Limbo Match reaches it through the dial: a player alone at midnight can find a category that fits them to within a hair, instantly.

But there is always an opponent, and often a real one

It would be wrong to file this game under solitaire. Three different opponents are available, and the player chooses which.

Opponent	How it works	Strength
The record holder	Every category holds a record — a specific number, set by a named player. 220,407 standing challenges.	Any level you choose
Friend Challenge	Share a result and its category; a friend plays the same shape and tries to beat it. A record share invites the reader to watch the replay first, then challenge it.	Someone you know
Tournaments (GLTX)	Real players, one event, one board shape, live standings.	The field that turned up

So the honest comparison with chess is not "solitaire against opponent". Chess requires a live opponent of matching strength at the moment you want to play. Limbo Match offers an opponent of *any* strength at any hour — because a record is an opponent who never declines, never has to be found, and can be watched before being challenged — and adds live challenges and tournaments on top when a player wants them. That is more choice, not less.

There is a version of this that matters particularly inside a family. A parent's record on a small category is an opponent of exactly the right strength for a child — present at any hour, patient, never condescending, and beatable with enough work. The child is not being let off, and the parent is not having to lose on purpose. When the record does fall, it falls to a real number that was genuinely there to be taken. Very few games can offer a five-year-old a fair contest against an adult without either bending the rules or bruising somebody, and this one does it with no mechanism beyond keeping the score.

20 · Across the tiers: one game, every age

This is the section the document is named for. The three shipped tiers are not three difficulty settings, and they are not a children's version, a demo and a real version. They are three doors into the same building, arranged so that a family can walk in through different doors and end up in the same room.

Start with what is actually shipped, because the ranges were chosen and they overlap on purpose. KIDS opens on 3 to 7 pairs; LITE is playable from 3 to 13; PRO runs the full 2 to 21.

	KIDS	LITE	PRO
Board sizes	3–7 pairs	3–13 pairs	2–21 pairs
Chips	Always iconic — pictures, which are numbers	Iconic or numbered	Iconic to 13, numbered to 21
Category shapes	Record shapes only, one loop mode	Record shapes only	The full lattice — 220,407
Records	None kept	Submits to the world records	Records, Champs, GLM100, tournaments
Controls	New · Reset · Timer	The full game	The full game plus the worth dial
Skill tiers served	1–2	1–4	1–6

What KIDS removes, and what it refuses to remove

KIDS takes away the things a young child cannot yet use, and nothing else. There is no record submission, so there is no competitive pressure and no way to fail publicly. The board is small. The loop mode is fixed. What it does **not** do is change the game: the chips, the dial, the flip limit, the scoring and the invisible board are exactly the ones a record holder plays with.

Two details in KIDS are worth singling out, because they show the design thinking about a real household rather than an abstract user.

Detail	Why it matters
Chips are always iconic	A picture *is* a number here. A child who cannot yet read numerals plays the full game, not a reduced one — and moves to numerals later without learning anything new.
RESET is behind an adult gate	Clearing the board needs a clock reading or a rotating fill-in-the-blank — Beethoven, Mozart, da Vinci, a capital city, a Latin tag — that an adult finishes and a young child does not. A fresh one each time, so it cannot be memorised.

That second one is a small piece of design with a large assumption behind it: that a grown-up is somewhere nearby. It is a family feature, not a security feature — and it is the clearest signal in the codebase that the game expects more than one generation in the room.

The timer belongs in the same category. It is a control the child can see and use, not a clock the game imposes: turning it off removes every trace of time pressure and leaves a board that waits indefinitely. For a young player, or an older one who does not want to be hurried, that single switch decides whether the game feels like a puzzle or an examination — and the score is unaffected either way, because in the world game time only ever separates equal scores.

Why the ranges overlap

LITE opens on a random board of 3 to 7 pairs — **exactly the range KIDS covers** — and then runs on to 13. PRO starts below where both of them start and runs to 21. So a child moving up from KIDS lands on boards they already know, and an adult arriving at the public front door starts on the same boards a child does. Nobody is ever handed a step they cannot take, and nobody has to admit to using the easy version. The tiers do not sort people; they sort ambitions.

What never changes between the tiers — and why that is the whole trick

A five-year-old on three pairs and a record holder on twenty-one are playing the same game in a strict sense. Same reds staying face-down. Same one point per match and one per unused flip. Same loops in the shuffle. Same technique — follow the chain — working at both ends of the ladder. **Nothing learned at the bottom has to be unlearned higher up**, and nothing at the top is a different game wearing the same name.

This is rarer than it sounds. The usual way to make a game work for a wide age range is to handicap it: give the child more time, fewer cards, a bigger clue, a head start. Every one of those tells the child they are being helped, and tells the adult they are not really playing. Limbo Match does something harder. It keeps the rules identical and moves the *board* instead — and because difficulty is a fine dial rather than three coarse settings, the board can be moved to fit anybody exactly.

The scoring then does the rest, quietly. A grandparent on seven pairs and a grandchild on four are not comparable in raw points, and the game never pretends they are — but both are measured against their own board's jackpot, both can see a personal best rise, and both are improving at the same thing. That is what a multigenerational game needs: not a level playing field, which is impossible, but a *shared measure of getting better*, which is not.

The merit system finishes the argument

Everything in sections 15 to 18 pays off here. Because merit is measured against the shuffle a player actually faced — the gift subtracted, the luck capped — a modest board played beautifully can out-score a large board played loosely. The eleven-year-old who traces every loop on nine pairs is not filed below the adult who muddles through sixteen. No age bracket is needed, no junior category, no separate table. The arithmetic already treats them as what they are: two people, each measured against what they were dealt.

A player is not moved between tiers by a level system, and never told they have been promoted. They move when the room they are in stops being big enough — which is the only motivation that lasts, at any age.

Two people at one table

The strongest test of a multigenerational game is not whether two ages can both play it. It is whether they can play it *together* and both take it seriously. Limbo Match passes that test in three separate ways, and none of them required a special mode to be built.

Together, how	What each side gets
Same board, taking turns	One shuffle, two attempts, two scores out of the same jackpot. The comparison is exact and needs no handicap.
Different boards, same measure	A child on four pairs and an adult on twelve each score against their own ceiling. Both can see a personal best move.
One plays, one watches	The board reveals nothing, so watching is real spectating — and the watcher learns the technique by seeing it used.

The third row is a consequence of the very first design choice in this document. Because matched reds stay face-down, an onlooker cannot read the state of the game from the board and cannot accidentally coach. What they *can* see is the chain being followed — the one thing worth teaching. A game that hides its progress but shows its method is close to ideal for a parent sitting beside a child.

What would break this, and has been avoided

It is worth naming the failure modes, because each was available and each was declined.

The easy path	Why it would have failed	What was done instead
A junior category	Splits the community and tells a child their result is not the real one	One measure, absolute, computed per shuffle
Age or skill handicaps	Both sides know the result is adjusted, so neither trusts it	The board moves, the rules never do
A simplified KIDS ruleset	Everything learned early would have to be unlearned later	KIDS removes tools, not rules
Levels and unlocks	Progress becomes the game's decision, not the player's	220,407 categories, all open, all the time

The pattern is consistent, and it is the reason the all-ages claim holds up under pressure rather than only in marketing. Nothing was made easier for anybody. The game was made *adjustable* instead, and then measured honestly enough that a small board played well is visibly worth more than a large board played badly.

Using the three tiers in practice

The ranges above are shipped facts; what follows is how they fit real situations. Four are worth spelling out, because in each one the tier choice and the board size do almost all of the work and nothing else has to be configured.

Situation	Tier	Boards	Why
A child's first afternoon	KIDS	3, then 4	Iconic chips, no clock pressure, no record to lose. Stop while it is still fun; the small ladder makes that easy.
A family evening, mixed ages	KIDS and LITE side by side	3–7 each	Everyone is inside the same range. Take turns on one shuffle for an exact comparison, or run separate boards and compare personal bests.
A classroom or club	LITE	5–9	Records are real and shared, the shapes are the canonical ones, and every result is comparable because the categories are fixed.
An adult keeping sharp	LITE, then PRO	7–13, then up	Start where the ladder is well trodden; move to PRO when the wish is for a particular shape rather than a bigger board.

Two practical notes that fall out of the mathematics rather than out of taste. First, the right board for a new player is **smaller than it looks**: at three and four pairs the whole ladder of technique is already present — loops, the flip economy, the endgame clover — and it is present in a form a beginner can actually see working. Nothing is gained by starting bigger, and confidence is lost. Second, the honest way to raise the difficulty is one dial at a time. PRO's worth-stepped Harder and Easier controls exist exactly for this: each press moves one visible dial to the next notch of worth, so the step is real but never a cliff.

And a caution, since this document should be useful rather than only admiring. The game is a game. It asks for attention and it measures attention, and it can be a fine thing to share across generations — but no claim is made here about training, therapy or cognitive benefit beyond the plain fact that people get better at what they practise. The interesting claim in this section is narrower and firmer: that one unmodified rule set can be genuinely worth playing at five, at forty and at eighty. That much the design demonstrably delivers.

A final observation about the shape of the three tiers, which only becomes visible once they are set side by side. The ranges do not merely overlap — each one contains the whole of the next one's starting ground. KIDS covers 3 to 7; LITE opens inside that range and extends it; PRO starts below LITE and extends further still. There is no point on the ladder where a player has to leave familiar ground to reach the next tier, and no tier that begins where another stopped. That is a deliberate arrangement, and it is the structural reason the phrase "across the tiers" describes a continuous climb rather than three separate products sharing a name.

21 · What is unreachable, and why that is the point

The jackpot requires every pair to be taken blind. The probability of that on a random board is 1 ÷ P! — for eight pairs, one in 40,320; for twenty-one pairs, one in 51 quintillion. So the top of every category is sealed off from skill alone. The best simulated expert reaches 78% of the ceiling. The last 22% needs the board to smile.

This is a feature, and a well-understood one. A game whose ceiling can be touched is a game that ends. Chess has no attainable perfect game; nor does golf; nor does music. Limbo Match keeps its summit permanently just out of reach, and keeps the gap between a player's current score and their reliably achievable score open at every level of play. That gap is what Grego felt as "almost infinite". It is not a feeling. It is a measurable property of the scoring rule.

Note how neatly this sits beside the merit system. The jackpot is unreachable, but the *ceiling* — the most that memory alone can take — is reachable, and reaching it is exactly what a top GLMP means. The game therefore has two summits: one that skill can stand on, and one that only luck can visit.

It is worth pausing on how rare that arrangement is. Most games with a luck component let the lucky result stand as the achievement — the record book cannot tell a brilliant hand from a kind one. Most games without luck have no summit above skill at all, so the very best play is also the best possible result and the game is, in principle, finished. Limbo Match splits the two apart and keeps both: a luck summit that nobody can train for, and a skill summit that is

precisely defined, precisely measurable, and genuinely hard. A player can spend years climbing the second while the first stays where it is, decorative and out of reach.

For a beginner the arrangement is kinder than it sounds. On a small board the skill summit is genuinely reachable: at three or four pairs a careful player can take everything memory allows, and will know they did. That is a complete, satisfying result available on the first afternoon — and it is the same kind of result a record holder is chasing at twenty-one pairs, not a consolation version of it. The unreachable summit stays unreachable for both of them, which is what keeps the game from ever being finished; the reachable one moves up as the board grows, which is what keeps it worth returning to.

The two summits also explain a thing players notice early and often misread. A run of extraordinary scores does not mean a sudden jump in ability, and a run of ordinary ones does not mean a slump: both are the luck summit moving in and out of reach while the skill summit stays exactly where it was. The merit number is the one that will not lie about this, because it has the luck capped out of it by construction. A player who wants to know whether they are actually improving should watch that number and ignore the scoreboard for a month.

22 · How it compares

	Concentration	Canasta	Sudoku	Chess	Limbo Match
Board gets easier as you win	Yes	Yes	Yes	—	**No**
Difficulty dial	None	None	Coarse	Via opponent	**220,407 steps**
Opponent	Optional	Required, 2–4	None	Required	**Optional — always on record**
Luck present	Yes	Yes, heavy	No	Almost none	**Yes, and measured**
Luck separated from merit	No	No	—	—	**Yes — floor and ceiling**
Skill layers	1–2	2–3	2–3	Many	**Six**
Rating system	None	None	Time	Elo / Glicko	**GLMP / GLM100**
Ceiling reachable by skill	Yes	No	Yes	No	**No**
Speed as a ranked axis	No	No	Sometimes	Clock only	**Always**
Same game for child and expert	Partly	Partly	No	Yes	**Yes**

Canasta earns its place in this table because it is the closest thing to Limbo Match in spirit: a card game of real skill in which luck is undeniably present, played for score rather than checkmate. The difference is what each does about the luck. Canasta lets it stand — a good hand is a good hand, and the scoring cannot tell the two apart. Limbo Match measures the deal, subtracts what it gave away and caps what it could never have given, so the number that ranks a player is what the player added. That is the whole argument of sections 15 to 18, compressed into one row of a table.

The two columns that matter most are the last two. On the dimensions that make chess deep — an unreachable ceiling, many skill layers, one board for all abilities — Limbo Match scores the same. On the dimensions where chess is awkward — needing a live opponent, needing a rating pool to find one — it does better, without giving up the opponent at all.

One row in that table deserves defending, because it looks like special pleading. Canasta is listed as having luck present and unmeasured, and a Canasta player might fairly say that skill shows over a long evening regardless. That is true, and it is exactly the point: the luck averages out over many hands, so the game needs volume before it can tell a good player from a lucky one. Limbo Match does not need volume, because it separates the two *within a single game*, by subtracting what the deal gave and capping what it could never have given. Both approaches work. One of them takes an evening and the other takes one shuffle, and only one of them can rank a child who plays twice a month.

23 · Three implications

The depth is invisible to a newcomer. A stranger sees a memory game. They cannot see the loops, the flip economy or the clover window. Chess solves this by letting people watch masters play. Limbo Match already has the tool — the record log stores full action recordings, and the share text already invites a reader to watch a record before challenging it. Replay is not a nice extra; it is the discovery mechanism.

The online game and the supervised tournament are two different products, and both are needed. An online rating is real and earned; a title is won under supervision. Chess proves that the two coexist happily, and that trying to make one do both damages both.

The mathematics is an asset, not a footnote. The flip ladder tracks the Golomb–Dickman constant. The tracing result is the 100 prisoners problem, made playable. The jackpot is 1 ÷ P!. The merit system measures a player against a computed floor and ceiling of the very deal they faced. These are not decorations on a game — they are reasons to take it seriously, and the flip sequence is already registered in Grego's own name as OEIS A389262.

24 · Conclusion

Grego's observation was correct, and the reason is simpler and stronger than it first appears. One rule — matched reds stay face-down — removes every form of help the game could have given the player. From that one rule follows a flat visible difficulty, a rising real difficulty, a currency in which every flip is priced, a hidden permutation whose loops reward a memory-free technique, a lottery button that becomes a certainty in trained hands, a forgiveness rule that becomes an opening theory, and a rotation that can rewrite the board without spending anything.

The ladder that sits on top of all this is calibrated so that perfect play succeeds roughly half the time, from five pairs to twenty-one. That is a knife edge, held for twenty rungs. And the ranking built above *that* refuses to measure a player against anybody else at all: it measures them against the deal they were given, subtracting the gift and capping the luck, which is a harder thing to build and a fairer thing to hold.

The result is a game that a child can play on their first afternoon and an expert can still be improving at years later, on a board that looks identical to both.

And that, finally, is what the title means. The tiers of this game are not only KIDS, LITE and PRO. They are the tiers of a mind: sight, recall, elimination, tracing, economy, tempo — and a person climbs them at their own pace, at whatever age they happen to be. The three products exist so that the climb can start anywhere and never has to stop. A child of five and a grandparent of eighty can be on the same ladder, on different rungs, playing rules that do not differ by a single word between them. Very few games can say that, and none of the ones that can got there by accident.

That is what chess is. It is not a common thing to have built.

A closing word about what this analysis could and could not settle. It could measure: the loops, the tail, the calibration, the clover, the pin, the worth of a category, the shape of the three tiers. Every one of those turned out to support the case rather than complicate it, which is unusual and worth saying plainly. What it could not measure is the part that actually decides whether a game lasts — whether people want to come back to it. No amount of combinatorics answers that, and no document should pretend otherwise.

But there is a reasonable inference to draw. Games that hold people for decades tend to share the properties catalogued here: a summit out of reach, a technique that repays study, a fine gradation of difficulty, an honest measure of who played well, and a single set of rules that does not care how old you are. Limbo Match has all five, by construction rather than by luck. Whether it finds its players is a question for the world; whether it deserves them is a question this document can answer, and the answer is yes.

25 · What would make this document wrong

An analysis that cannot be contradicted is not worth much. So each main claim is set out below with the evidence that would overturn it — and, more usefully, with what is actually known today and how much confidence that supports. Two of the five are settled, one is strongly supported by an experiment run for this document, one turned out to need its wording corrected, and one is simply open.

The claim	What would overturn it	Where it stands
The depth comes from matched reds staying face-down	A version that removed spent reds from the dial and played just as deep	**Tested here.** A no-memory player scores 15.7 of 48 under the shipped rule and 25.2 if spent reds are removed — the rule alone is worth a 60% swing in what the board gives away free. Strongly supported.
Tracing does not raise the average, only the tail	Play data showing self-taught tracers have higher mean scores, not just higher bests	**Settled.** The two marginal distributions are provably identical, and simulation agrees at every flip limit. One caveat: real tracers also remember, so observed means will differ — the claim holds memory constant.
The record ladder is calibrated near a coin flip	An arithmetic error in the cycle enumeration	**Settled.** Exact rational enumeration, reproducible in under a second by anyone who wants to check it.
Merit measures the player, not the deal	Merit tracking how kind the shuffle was, rather than how well it was played	**Corrected — see below.** Measured across 60,000 shuffles: a fixed-skill player's raw score is independent of the shuffle (r = +0.00), but their GLMP moves almost exactly opposite to the floor (r = –0.99).
The game works across all ages	Usage showing KIDS players never move up, or that older players abandon the larger boards	**Open.** No play data exists. This is a design argument only, and it is offered as reasoning rather than evidence.

The one that needed correcting

"Merit measures the player, not the deal" is too strong, and the measurement says so. Hold skill fixed and vary only the shuffle: the raw score barely moves — a player with memory needs about the same number of flips whatever the loops look like — while GLMP swings from near nothing on a generous board to nearly the whole score on a bare one.

That is not a fault. It is the anti-farming property working exactly as intended, and now measured: **grinding kind shuffles earns almost no merit**, and no eligibility rule was needed to make it so. But it does mean GLMP is a property of *a game*, not a stable trait of a person. What converts it into a measure of a player is the rule from section 17 — rank by the **best single** GLMP. A high standing therefore means something precise: not "plays well on average", but "once met a demanding shuffle and took nearly everything it had to give". That is a harder and more interesting thing to have done, and it is what the leaderboard is actually reporting.

Two consequences follow that a player would want to know. Chasing records on easy categories is not merely unrewarding, it is close to pointless — the merit simply is not there to be won. And a disappointing GLMP after a good-feeling game usually means the board was kind, not that the play was poor.

A second correction: something this document had simply missed

Section 11 now carries the **deliberate-limbo harvest** — clovering a returning blue for 1 + F rather than flipping it for F. The first edition did not mention it at all, in a section devoted to the limbo.

It was not unknown to the project. `limbomaths.pdf` names the tactic and states its mechanism and its cost precisely; the shipped ceiling model has been playing the line all along, with the branch that does it carrying the comment *"force it down: 0 now, 1+F later"*. So this is not a discovery. It is a document that examined a rule for several pages and failed to notice the strongest thing that can be done with it, while a sibling document had it written down.

The lesson is narrower and more useful than "check the maths": **read what the project has already concluded before analysing it afresh.** The gap here was between documents, not between the document and the game.

The one this revision had to correct

Section 15 previously claimed that worth was **exactly computable**, resting on the proposition that a deal's ceiling depends only on its loop shape. Both halves were wrong, and the correction is set out in full there.

What is worth recording here is *how* the error survived. It was tested — the document said so, and the test was real: three hundred blue orders, zero spread. But the test varied the wrong thing. It held the deal fixed and varied the blues, when the failure was in holding the shape fixed and varying the deal. **A verified claim is only as good as the variable the verification moved.**

It was not caught by re-reading the model. It was caught because a live player achieved a merit of 27 in a category whose theoretical maximum this document had put at 23. The system's own record log falsified the system's own bound, and it took eleven months of play for a game to land in the right category to do it.

That is the most useful thing in this revision, and it argues for something the original document did not say: **the standing records are themselves a test suite.** Every stored game is a claim about what is achievable, checkable against the model that says what should be achievable. Nobody had run that check until now.

Two kinds of claim, and different confidence in each

The failure conditions divide cleanly. The mathematical claims — the loops, the calibration, the clover values, the pin, the floor arithmetic — cannot be overturned by evidence at all, only by a mistake in the working, and the working is short enough to check. The claims about people are the opposite: predictions about behaviour that only play data can settle. This document is confident about the first kind and explicitly not confident about the second.

There is one more honest limit, and it is the largest. Everything here describes a game that is finished and correct. Nothing here says whether people will find it, or whether the first five minutes are inviting enough to reach the depth catalogued above. A game can be excellent and unplayed. That is not a flaw in the design, and no analysis of the design can fix it — but it would be dishonest to end a document this favourable without naming the one thing it cannot speak to.

Appendix · How the numbers were produced

Nothing in this document was taken on trust from the game's own reporting. Two independent methods were used, and where both applied they agreed.

Exact combinatorics. The expected longest loop and the clean-sweep probabilities in section 7 were computed by full enumeration of cycle structures in rational arithmetic — not estimated. The count of permutations of P elements with no loop longer than m follows a standard recurrence; dividing by P! gives the probability exactly. The clover values in section 10 are closed-form.

Simulation of the shipped rules. Where a player's judgement is involved, no formula exists, so the rules were re-implemented and played: 40,000 games for each of the five players in section 8, 20,000 boards for each sweep and pin estimate. What the simulation does *not* model is a real human: it has either perfect memory or none. Real players sit between the two, which is precisely the space this document argues is deep.

Figures taken from the shipped system. The worth values, the floor and ceiling definitions, the band counts, the percentiles and the nineteen World Record rung worths in sections 15 to 17 are not this analysis's own calculations. They are read from the shipped engine and its reference implementation, and are quoted here as the system's own published numbers.

What this analysis does not claim. Four honest limits. First, the player models are idealised; a real beginner is worse than player A on a bad day and better on a good one. The *shape* of the ladder is the claim, not the exact means. Second, time was not simulated — every result here is about score, so the second axis of section 14 is described but not measured, and note that this is a gap in measurement, not a missing tradeoff: speed and accuracy here are two outputs of one cause. Third, the calibration finding in section 7 is an observation about what the ladder *does*; it is not a claim about how it was derived. Fourth, the originality verdict in section 18 is bounded by what I know of the literature; it is a considered opinion, not a search of every journal.

How to check any of this. Every claim in sections 5 to 8 can be re-derived from the rules alone, with no access to the game's code: shuffle a deck of P numbered cards, write down the loops, and the flip counts follow. The claims in sections 15 to 17 can be checked against the shipped engine, where the floor, the ceiling and the worth sweep are all implemented twice — once in the browser and once on the server — deliberately kept line-for-line identical so that the board a player sees and the record the server stores can never disagree. That duplication is not redundancy; it is the audit.

A word on what was hardest to pin down. Two things resisted. The first was the claim that tracing and random scanning have the same expected score: it looks wrong, and the first simulation was written twice before I accepted that the code was right and the intuition was not. The second was the all-ages argument, which is easy to assert and hard to evidence — the useful move was to stop reasoning about children and instead read what the three tiers actually ship: the board ranges, the always-on iconic chips, the absence of record submission in KIDS, and the adult gate on its reset. Those are facts, and they carry the argument better than any amount of sympathy for beginners.

The two experiments run for section 25. Both are simple and both are reproducible. The first removes matched reds from the dial and replays a memoryless player on the same shuffles, isolating the effect of the one design rule this document builds everything on. The second holds skill constant and varies only the deal: for each of 60,000 shuffles it computes the floor, runs a perfect-memory line to get the ceiling, runs an ordinary memory player to get a score, and correlates the resulting merit with the floor. Neither needed a new model — both reuse the players from section 8 and the floor formula from section 16.

The ceiling used in the second experiment is the same conservative one described in section 16: memory, tracing, elimination and the endgame clover, but no Pin and no deliberate limbo. A fuller ceiling would sit slightly higher, which would shift the merit values a little and the correlation not at all.

Sources consulted. Where a figure came from the live system rather than from a calculation made here, it came from one of these. Client-side files are named, because anyone can already read them in a browser. Server-side code is described by its role rather than its filename — naming a server file in a public document tells a reader nothing useful and a prober something useful, which is the wrong trade.

Source	What was taken from it
limboGLRS.js	The merit block: the GLMP formula, the floor and ceiling definitions, and the ranking rules
The server-side worth code	The twin of the browser computation, kept line-for-line identical with it
The floor/ceiling reference model	Grego's own implementation, which the shipped code follows
limboWorthLUT.json	The worth range 1–441, the Feast count of 220,407, and the percentile table behind the difficulty slider
limbomatch.html (PRO)	The World Cup pool, the category test, the GLM100 rules, and the Friend Challenge share
limbo.html (LITE)	The 3–13 board clamp, the opening 3–7 range, and the single loop mode
limbokids.html (KIDS)	The 3–7 range, the always-iconic chips, the absence of submission, and the adult gate on RESET
limbostrategy.pdf	The cover illustration of the two readers, Cyrus and Rowen
limbomaths.pdf	The deliberate-limbo harvest, named and defined there before this document reached it
OEIS A389262	The flip-limit sequence MF[P], registered in Grego's name

Nothing was taken from marketing copy, from the game's own explanatory text, or from any claim the system makes about itself. Where the code and the documentation disagreed, the code was taken as the truth — twice in this session that distinction mattered, and both times the code was the thing actually being played.

A word on what was hardest to pin down. Two things resisted. The first was the claim that tracing and random scanning have the same expected score: it looks wrong, and the first simulation was written twice before I accepted that the code was right and the intuition was not. The second was the all-ages argument, which is easy to assert and hard to evidence — the useful move was to stop reasoning about children and instead read what the three tiers actually ship: the board ranges, the always-on iconic chips, the absence of record submission in KIDS, and the adult gate on its reset. Those are facts, and they carry the argument better than any amount of sympathy for beginners.

One practical note for anyone re-running the figures. The exact calculations are cheap — the cycle enumeration for all twenty board sizes finishes in under a second in rational arithmetic, so there is no reason to approximate it. The simulations are the slow part, and the sample sizes here were chosen so that the standard error on every mean quoted is below 0.05 points. Where a figure is quoted to one decimal place, that decimal is real. Two figures are quoted rather than derived and should be treated as the system's own: the 220,407 category count and the nineteen rung worths.

The World Record ladder in full

P	F	Jackpot	Longest loop	Clean sweep	Worth
3	2	9	2.17	66.7%	8
4	3	16	2.79	75.0%	14
5	3	20	3.42	55.0%	16
6	4	30	4.04	63.3%	25
7	5	42	4.68	69.0%	36
8	5	48	5.30	56.5%	41
9	6	63	5.92	62.1%	55
10	6	70	6.55	52.1%	60
11	7	88	7.17	57.3%	77
12	8	108	7.80	61.5%	96
13	8	117	8.42	53.8%	103
14	9	140	9.05	57.7%	125
15	9	150	9.67	51.1%	133
16	10	176	10.30	54.8%	158
17	11	204	10.92	58.0%	185
18	11	216	11.55	52.5%	195
19	12	247	12.17	55.5%	225
20	12	260	12.79	50.5%	236
21	13	294	13.42	53.5%	269

"Longest loop" is the expected length of the longest loop on a freshly shuffled dial. "Clean sweep" is the exact probability that a perfect chain-follower clears every pair without one miss. "Worth" is the shipped category worth for the rung. P = 2 is not a record rung: it was retired from the World Record ladder, and is custom-record only.

Glossary

Term	Meaning	Term	Meaning
P	Pairs — blue chips, and reds on the dial	Floor	What a shuffle gives free to no memory at all
F	Flip limit — reds turnable in one search	Ceiling	The most perfect memory can take without gambling
L	Limbo allowance — early blues forgiven a miss	Room	Ceiling minus floor — what is at stake in a deal
Loop / trail	A closed chain of positions and values	Worth	The largest room a category can offer; 1 to 441
Jackpot	P × (F + 1) — reachable only by luck	GLMP	max(1, min(score, ceiling) − floor); a provably blind run uses blind bounds
MF[P]	The record flip limit — OEIS A389262	GLMR	A player's rank, from their best single GLMP
Feast	All 220,407 categories together	GLM100	The published top of the GLMR list

Questions this opens

Three that a further study could answer, and that this one deliberately did not. **What does the time axis look like?** Nothing here measures it, yet it is half the ranking, and the interesting question is not speed against accuracy but how quickly a player's fluency follows their understanding. This revision tried and failed to answer a narrower version of it — whether merit-heavy games run faster — and found the standing records too few and too unevenly spread across board sizes to say. It is the first question a larger body of play data would settle. **Where does a real human sit between the two simulated extremes?** The models have perfect memory or none; the interesting territory is between them, and only real play data can map it. **Does worth predict how much players enjoy a category?** The system prices difficulty honestly — whether players are drawn to the honest price is a separate and testable question.

What the analysis suggests building next

Three things follow from the findings rather than from opinion, and each is small.

Suggestion	Why this analysis points at it
Put replay in front of newcomers	The depth is invisible from outside, and the record log already stores full action recordings. Watching one traced game teaches more than any explanation.
A short strategy page for players	Sections 6, 10, 11 and 12 contain five rules nobody would guess — clover the last blue, take the free coin toss at two, spend the first blue scouting, pin a dead board, and harvest the limbo queue — and each is worth points immediately. The harvest is the one to lead with: it is worth a point per limbo, it is pure memory rather than luck, and section 11 shows a real player leaving six points and a title on the table for want of it.
Watch a real player's memory at work	Section 9 names three routes — rehearsal, visual, and chunked meaning — and the game engages all three. Which one carries a given player is unmeasured, and it is the one genuinely open question here.

None of the three requires a rule change, and none touches the mathematics. That is itself a conclusion worth stating: after twenty-four pages of examination, the recommendations are about showing the game to people, not about altering it.

Analysis by Claude (Anthropic) for Grego, 25 July 2026. All figures computed for this document unless marked otherwise: exact combinatorics where available, otherwise simulation of the shipped rules (40,000 games per player; 20,000 boards per pin and sweep estimate). Worth, floor, ceiling, GLMP, GLMR and GLM100 figures are read from the shipped engine. Limbo Match, GLRS, GLM100, GLTX and the flip sequence OEIS A389262 are the work of Grego (Földvári Gergely).

A note on method. Two habits were followed throughout. Nothing was asserted from memory where it could be computed instead, and nothing was computed where the shipped engine already holds the answer — in that case the engine's number is quoted and marked as such. Where the two overlapped, they were checked against each other. Every figure in this document can therefore be traced to one of three places: the rules of the game, an exact calculation reproducible from those rules, or a named file in the live system.

Provenance. This document was written in one session on 25 July 2026, at Grego's request, after a working day spent on the game's fan logic and on completing the retirement of the two-pair record category. The observation that prompted it was his own, made while playing. Nothing in it was drafted in advance, and every number in it was produced during the writing, from the rules and from the shipped engine.

Revision, 28 August 2026. This edition was not planned. It began with a game: Grego challenged a standing record in an eight-pair category, scored four points fewer than the holder in a little over half the time, and lost — by a single point of merit. Working out why turned up two things this document had got wrong and one it had never noticed.

The first was a published bound. The holder's game measures a merit of 27 in a category whose worth this document put at 23. A real player had exceeded a theoretical maximum, which meant the maximum was wrong. Section 15's claim that worth is exactly computable is withdrawn, and section 25 says how it survived: the original test was real, but it moved the wrong variable.

The second was an omission. The deliberate-limbo harvest — the strongest thing that can be done with the limbo — had been named and defined in the project's own maths paper, while this document, in a section devoted to the limbo, did not mention it. It is now section 11, credited where it belongs.

A third claim was written, tested and thrown away: that merit-heavy games are played faster. Thirty-five records across nine board sizes could not carry it. The attempt is in section 25 rather than quietly deleted, because a document that reports only the tests that worked is not reporting its method.

The game itself changed in the interval — merit can now take a category title from a higher score, which is section 17. And it changed again while this edition was being finished: GLMP now prices a provably blind run on elimination alone, documented in section 16. That correction is Grego's, not mine — I had folded it into an earlier decision where it did not belong, and he separated the two questions. As with the first edition, where I was wrong Grego corrected me; this time more than twice.