ENDPOINT DOMINANCE FOR A RESTARTING JOSEPHUS PROCESS: THE FRONTIER AT $n \geq 68$ AND THE POSITION OF THE INTERNAL MAXIMUM

KOREA SUPERINTELLIGENCE LABS

Abstract. The restarting Josephus process is the threshold recurrence given by $F_1(q) = 1$ and $F_k(q) = F_{k-1}(q) + \mathbf{1}_{\{q \bmod k < F_{k-1}(q)\}}$, whose value $F_n(q)$ is the initial position of the survivor in a line of n people counted out with step $q + 1$, the count restarting at the leftmost survivor after every deletion. It depends on q only through $q \bmod L_n$, $L_n = \mathrm{lcm}(1, \dots, n)$, so the fiber sizes $N_{n,j} = \#\{q \bmod L_n : F_n(q) = j\}$ are well defined; both endpoint fibers have size exactly $\varphi(L_n)$. Chen asks whether every internal fiber is strictly smaller — whether the two endpoints are the unique global maxima — and verifies this for $4 \leq n \leq 49$. We compute the exact fiber-size vectors for $50 \leq n \leq 66$, reproducing all 49 published rows digit for digit along the way, and find endpoint dominance at every level; since 67 is prime, one formally verified propagation step carries the verification to $n = 67$, so a least counterexample must satisfy $n \geq 68$. We then isolate a regularity that the computation exposes and the published account does not record: in all 64 computed rows the internal maximum sits at $j = 2$. We prove that this positional statement alone implies endpoint dominance, with an explicit slack $\gamma_n > 0$ obtained by carrying a quantitative term through the induction that proves $N_{n,2} < \varphi(L_n)$. Finally we extend the majorization defect $D_{n,k}$ past the published range: the sign pattern persists at every composite $n \leq 66$, the record slack $\min D_{n,k}/L_n$ is beaten by a factor 2.146 at $n = 65$, and $D_{n,3}$ coincides at every level checked with the slack of the composite lift $N_{n,2} \leq \rho_n N_{n-1,2}$, so that the route proposed in the source reduces to the same statement about where the internal maximum sits. Every statement valid for general n is machine-checked in Lean 4; the exact vectors for $n \geq 15$ are the output of an independent program in C, and we say throughout which is which.

1. Introduction

The process. Fix an integer $q \geq 0$ and count out a line of people with step $m = q + 1$, but let the count restart at the current leftmost survivor after *every* deletion, the survivors keeping their left-to-right order. At population size $k \geq 2$ the deleted rank is $1 + (q \bmod k)$, and reverse insertion of the new last person turns the rule into the threshold recurrence

$$F_1(q) = 1, \qquad F_k(q) = F_{k-1}(q) + \mathbf{1}_{\{\, q \bmod k \,<\, F_{k-1}(q)\,\}} \qquad (k \geq 2), \tag{1}$$

in which $F_n(q) \in \{1, \dots, n\}$ is the initial position of the survivor in a line of n. This is the *restarting* Josephus process of Chen [1]; it is not the classical circular process, in which the count continues from the next survivor. For the classical problem see [3, Section 1.3] and [4]; the diagonal case $m = n$ of the restarting rule is recorded as OEIS `A128982` [5], so the restart convention itself is not new, but the two-parameter arithmetic of (1) was taken up in [1].

Writing

$$L_n = \mathrm{lcm}(1, 2, \dots, n),$$

every level $k \leq n$ divides L_n, so F_n depends on q only through $q \bmod L_n$ and the *fiber-size vector*

$$N_{n,j} = \#\{\, q \bmod L_n \;:\; F_n(q) = j \,\} \qquad (1 \leq j \leq n)$$

is well defined. It sums to L_n and is a palindrome, $N_{n,j} = N_{n,n+1-j}$, because $q \mapsto L_n - 1 - q$ reverses the survivor position. Its two endpoint entries are not merely large but exactly computable:

$$F_n(q) = 1 \iff \gcd(q, L_n) = 1, \qquad \text{hence} \qquad N_{n,1} = N_{n,n} = \varphi(L_n). \tag{2}$$

The question. Since the endpoint value is exactly $\varphi(L_n)$, the natural question is whether an internal survivor position can be at least as likely.

Problem 1.1 ([1, Open Problem 6.1])**.** For every $n \geq 4$ and every $1 < j < n$, is $N_{n,j} < \varphi(L_n)$? Equivalently, are the two endpoints the unique global maxima of $(N_{n,1}, \ldots, N_{n,n})$?

We say that *endpoint dominance holds at level* n when the inequality of Problem 1.1 holds for every internal j. The hypothesis $n \geq 4$ is necessary: $(N_{3,1}, N_{3,2}, N_{3,3}) = (2, 2, 2)$. The problem is not a statement about monotonicity — already $N_{12,5} > N_{12,4}$ — but only about the two maxima.

What [1] settles is the following. At a prime level p the new modulus is coprime to L_{p-1}, so the Chinese remainder theorem gives the transition formula $N_{p,j} = (p - j)N_{p-1,j} + (j - 1)N_{p-1,j-1}$ [1, Proposition 6.2] and hence propagation of endpoint dominance through prime levels [1, Corollary 6.3]; the two positions adjacent to the endpoints obey $N_{n,2} = N_{n,n-1} < \varphi(L_n)$ for every $n \geq 4$ [1, Theorem 6.4]; and consequently a least counterexample would be composite with offending position $3 \leq j \leq n - 2$ [1, Corollary 6.5]. An exact computation verifies endpoint dominance for $4 \leq n \leq 49$ [1, Proposition 6.9], which leaves the frontier

$$n \geq 50, \qquad n \text{ composite}, \qquad 3 \leq j \leq n - 2. \tag{3}$$

The source also records a majorization defect $D_{n,k}$ (Section 6 below) whose nonnegativity at $k = 3$ would settle Problem 1.1, and reports the sign pattern of $D_{n,k}$ over its computed range together with the smallest observed value of $D_{n,k}/L_n$.

What is settled here. This paper moves the frontier and, more to the point, changes what the remaining question is about.

The frontier (Section 4). An independently written program computes the exact fiber-size vectors for $50 \leq n \leq 66$; its output for $n \leq 49$ agrees with the published table digit for digit, and its output for $n \leq 22$ agrees with a direct enumeration. Endpoint dominance holds at every one of the new levels. Level 67 is prime, so the prime-propagation step — here a formally verified theorem, valid for every n — carries the verification one level further with no computation at all. A least counterexample must therefore satisfy $n \geq 68$.

The position of the internal maximum (Section 5). In all 64 computed rows with $n \geq 3$ the internal maximum of the fiber-size vector is attained at $j = 2$ (and, by the palindrome, at $j = n - 1$). The source's own program prints this quantity, but its text never states where the maximum sits. We prove that this one positional statement is enough: if $N_{n,j} \leq N_{n,2}$ for every internal j, then endpoint dominance holds at n, with the explicit slack γ_n of Theorem 3.4. Problem 1.1 is therefore entirely a question about the location of a single maximum. We do not prove that location for general n; we prove that nothing else is missing.

A quantitative nearest-internal theorem (Section 3). The induction of [1, Theorem 6.4] yields $N_{n,2} < \varphi(L_n)$ with no slack. Carrying a slack term through the same induction gives $N_{n,2} + \gamma_n \leq \varphi(L_n)$ with $\gamma_n > 0$ explicit, $\gamma_n = (L_n/6)\prod_{5 \leq p \leq n}(1 - 2/p)$. This is a routine strengthening of a published theorem and is claimed as nothing more; it is recorded because it is what makes the reduction of Section 5 quantitative.

The majorization defect (Section 6). The sign pattern of $D_{n,k}$ reported in [1] for composite $4 \leq n \leq 49$ continues at every composite $n \leq 66$. The smallest ratio $D_{n,k}/L_n$ over the extended range is attained at $(n, k) = (65, 3)$ and is smaller than the published record at $(39, 3)$ by a factor 2.146, so the margin in the finite evidence shrinks appreciably just past the published range. Moreover, at every composite level checked,

$$D_{n,3} = \rho_n N_{n-1,2} - N_{n,2}, \qquad \rho_n = L_n/L_{n-1},$$

which is exactly the slack in the composite lift $N_{n,2} \leq \rho_n N_{n-1,2}$ of [1, Theorem 6.4]. The route "prove $D_{n,3} \geq 0$ at every composite level", proposed in [1] as one way to settle Problem 1.1, thus reduces — on any range where the internal maxima sit at $j = 2$ — to a statement already proved, and the whole content returns to the position of the maximum.

Provenance of the data. The fiber-size vectors of [1] are public: contrary to its reproducibility appendix, which points at an unlisted view-only cloud link that the appendix itself describes as not providing permanent scholarly access, the arXiv e-print of [1] carries an ancillary directory holding the generator source, the checker, and the complete exact output for $1 \le n \le 49$ [2]. That output is what our program was checked against. (Checked on 2026-08-28: `2608.09092` has a single version, v1, and the three files of [2] are both named on its arXiv ancillary listing and present in its e-print.) A search of the OEIS on 2026-08-28 found neither the vector $(N_{12,1}, \dots, N_{12,12})$ nor the sequence $(N_{n,2})_{n\ge 4} = 2, 10, 8, 56, 99, 255, 240, 2736, 2538, \dots,$ so the fiber sizes appear not to be tabulated elsewhere.

2. Preliminaries

Throughout, $n \ge 1$ and $q \ge 0$ are integers, F_n is defined by (1), $L_n = \mathrm{lcm}(1, \dots, n)$ with $L_0 = 1$, φ is Euler's totient, and

$$N_{n,j} = \#\{\, 0 \le q < L_n \;:\; F_n(q) = j \,\}, \qquad \rho_n = \frac{L_n}{L_{n-1}} \quad (n \ge 1).$$

We set $N_{n,j} = 0$ for $j = 0$ and for $j > n$, which is consistent with the definition.

Proposition 2.1 (the recurrence and its period)**.** *For all $n \ge 1$ and $q \ge 0$ we have $1 \le F_n(q) \le n$; every level k with $1 \le k \le n$ divides L_n; and $F_n(q \bmod M) = F_n(q)$ for every multiple M of L_n. In particular F_n is periodic with period L_n, and $N_{n,j} = 0$ unless $1 \le j \le n$.*

Proof sketch. The bounds and the divisibility are inductions on n using (1). For the period, if $L_n \mid M$ then $k \mid M$ for every level $k \le n$, so $(q \bmod M) \bmod k = q \bmod k$ and each step of (1) reads the same threshold. □

Proposition 2.2 (the endpoint fibers)**.** *For all n, q,*

$$F_n(q) = 1 \iff \big(\forall k,\ 2 \le k \le n \Rightarrow k \nmid q\big) \iff \gcd(q, L_n) = 1,$$

and consequently $N_{n,1} = N_{n,n} = \varphi(L_n)$ for $n \ge 1$.

Proof sketch. The first equivalence is an induction on n: the state stays at 1 at level k exactly when $q \bmod k \ne 0$. For the second, a common prime factor of q and L_n is at most n, because every prime dividing L_n is; conversely a level $k \le n$ dividing q divides $\gcd(L_n, q)$. The count of $q < L_n$ coprime to L_n is $\varphi(L_n)$, which gives $N_{n,1}$; the value of $N_{n,n}$ then follows from Proposition 2.3. □

Proposition 2.3 (reflection)**.** *Let $n \ge 1$, let L be a multiple of L_n and let $0 \le q < L$. Then*

$$F_n(q) + F_n(L - 1 - q) = n + 1,$$

and therefore $N_{n,n+1-j} = N_{n,j}$ for every j.

Proof sketch. Induction on n, the step being the identity $(L - 1 - q) \bmod k = k - 1 - (q \bmod k)$ for every divisor k of L, which turns the threshold test at level k into its complement. The palindrome follows because $q \mapsto L_n - 1 - q$ is an involution of $[0, L_n)$ carrying the fiber over j bijectively onto the fiber over $n + 1 - j$. □

Proposition 2.4 (counting infrastructure)**.** *Let $n \ge 1$.*

(i) $\#\{0 \le q < cL_n : F_n(q) = j\} = c\, N_{n,j}$ *for every $c \ge 0$ and every j.*
(ii) *The primes dividing L_n are exactly the primes $p \le n$; if $n+1$ is not prime then L_{n+1} and L_n have the same prime divisors.*
(iii) *If $n + 1$ is not prime then $\varphi(L_{n+1})\, L_n = \varphi(L_n)\, L_{n+1}$; that is, the totient of the period grows by exactly the factor ρ_{n+1}.*
(iv) $\sum_{j=1}^{n} N_{n,j} = L_n$.

Proof sketch. (i) is periodicity plus a disjoint decomposition of $[0, cL_n)$ into c periods. (ii) is the definition of L_n together with the fact that a prime dividing a least common multiple divides one of its arguments. (iii) follows from $\varphi(m)\prod_{p|m} p = m\prod_{p|m}(p-1)$ applied to L_{n+1} and L_n, whose sets of prime divisors agree by (ii). (iv) is Proposition 2.1 and a fiberwise count of $[0, L_n)$. ☐

Direct enumeration of a full period is feasible in exact arithmetic while L_n is small; $L_{14} = 360\,360$.

Proposition 2.5 (the small vectors)**.** *The fiber-size vectors $(N_{n,1}, \ldots, N_{n,n})$ for $3 \le n \le 14$ are*

$$\begin{aligned}
n = 3 &= (2, 2, 2)\\
n = 4 &= (4, 2, 2, 4)\\
n = 5 &= (16, 10, 8, 10, 16)\\
n = 6 &= (16, 8, 6, 6, 8, 16)\\
n = 7 &= (96, 56, 40, 36, 40, 56, 96)\\
n = 8 &= (192, 99, 73, 56, 56, 73, 99, 192)\\
n = 9 &= (576, 255, 188, 165, 152, 165, 188, 255, 576)\\
n = 10 &= (576, 240, 176, 150, 118, 118, 150, 176, 240, 576)\\
n = 11 &= (5760, 2736, 1888, 1578, 1308, 1180, 1308, 1578, 1888, 2736, 5760)\\
n = 12 &= (5760, 2538, 1678, 1135, 1413, 1336, 1336, 1413, 1135, 1678, 2538, 5760)\\
n = 13 &= (69120, 33678, 21856, 15249, 15844, 16417, 16032,\\
&\qquad 16417, 15844, 15249, 21856, 33678, 69120)\\
n = 14 &= (69120, 31926, 21460, 14208, 14365, 14470, 14631,\\
&\qquad 14631, 14470, 14365, 14208, 21460, 31926, 69120)
\end{aligned}$$

and endpoint dominance holds at every level $4 \le n \le 14$.

Proof. Exhaustive evaluation of (1) at every $q \in [0, L_n)$, for each of the twelve levels, inside the proof assistant; the cost at level n is n passes over L_n values. The row $n = 12$ is the vector displayed as equation (40) of [1], and all twelve rows agree with the published table [2] digit for digit. ☐

The next proposition records the boundary of the question: what fails just outside it, and that its two sides are attained values rather than bounds.

Proposition 2.6 (negative controls)**.**
- (i) *Endpoint dominance fails at $n = 3$, so the hypothesis $n \ge 4$ in Problem 1.1 is load-bearing and no induction can start earlier.*
- (ii) *$N_{12,4} < N_{12,5}$: the fiber sizes do not decrease monotonically inward from an endpoint, so only the statement about the maxima can survive.*
- (iii) *$\varphi(L_5) < 2N_{5,2}$: no bound of the shape $2N_{n,j} \le \varphi(L_n)$ can hold at internal j.*
- (iv) *$N_{12,1} = \varphi(L_{12}) = 5760$ and $N_{12,2} = 2538$: both sides of the inequality of Problem 1.1 are attained values.*

Proof. Each item is a finite computation from (1) over one period, at $n \in \{3, 5, 12\}$; see Proposition 2.5. Item (ii) is the remark made in [1] just after Problem 1.1, here checked. ☐

3. Transitions, the composite lift, and an explicit gap

This section is the Section 6 apparatus of [1], proved for every n, together with one quantitative strengthening. Nothing in Propositions 3.1–3.3 or Corollary 3.6 is claimed as new mathematics; they are stated because the results of Sections 4–6 rest on them and because the slack of Theorem 3.4 has to be carried through their proofs.

Proposition 3.1 (prime transition; [1, Proposition 6.2])**.** *Let $p = m + 1$ be prime with $m \ge 1$. Then $L_p = pL_{p-1}$, and for $1 \le j \le p$*

$$N_{p,j} = (p - j)N_{p-1,j} + (j - 1)N_{p-1,j-1}. \tag{4}$$

Moreover $\varphi(L_p) = (p - 1)\varphi(L_{p-1})$.

Proof sketch. Since $p \nmid L_{p-1}$ we have $L_p = L_{p-1}p$ and $\gcd(p, L_{p-1}) = 1$. The proof given here makes the Chinese remainder step explicit in three lemmas: a shift by a step coprime to L permutes $[0, L)$ and hence preserves any count; consequently each residue class c modulo p meets the level-$(p-1)$ fiber over i in exactly $N_{p-1,i}$ points of $[0, L_{p-1}p)$; and summing over a set S of residues gives $\#S \cdot N_{p-1,i}$. Splitting the level-p fiber over j into the classes that stay at j (the $p-j$ residues $c \geq j$) and those that rise from $j-1$ (the $j-1$ residues $c < j-1$) gives (4). The totient identity is multiplicativity, $\varphi(L_{p-1}p) = \varphi(L_{p-1})(p-1)$. □

Proposition 3.2 (prime propagation; [1, Corollary 6.3])**.** *Let $m \geq 4$ with $p = m+1$ prime. If endpoint dominance holds at level m, then it holds at level p. (Under endpoint dominance at level n, every fiber satisfies $N_{n,i} \leq \varphi(L_n)$, including the endpoint and out-of-range values.)*

Proof sketch. For internal j at level p, at least one of the indices j, $j-1$ is internal at level m, so one of $N_{m,j}$, $N_{m,j-1}$ is strictly below $\varphi(L_m)$ while the other is at most $\varphi(L_m)$. Since the coefficients in (4) sum to m and $\varphi(L_p) = m\varphi(L_m)$, the strict inequality is preserved. □

Proposition 3.3 (the composite lift; from the proof of [1, Theorem 6.4])**.** *Let $n \geq 4$ with $n+1$ not prime. Then*

$$N_{n+1,2} \leq \rho_{n+1}N_{n,2} \qquad \textit{and} \qquad \varphi(L_{n+1}) = \rho_{n+1}\varphi(L_n).$$

Proof sketch. Every class modulo L_n has ρ_{n+1} lifts modulo L_{n+1}, so it suffices to show that the level-$(n+1)$ fiber over 2 is contained in the level-n fiber over 2: a final state 2 arises either by staying at 2, or by rising from state 1, and the latter forces $(n+1) \mid q$, whereas state 1 forces $\gcd(q, L_n) = 1$ — impossible, since the least prime factor of the composite $n+1$ is at most n and divides both. The totient identity is Proposition 2.4(iii). □

The induction in [1, Theorem 6.4] combines Propositions 3.1 and 3.3 to give $N_{n,2} < \varphi(L_n)$ with no quantitative slack. The same induction carries a slack term. Define

$$\gamma_4 = 2, \qquad \gamma_n = \begin{cases} (n-2)\,\gamma_{n-1}, & n \text{ prime,} \\ \rho_n\,\gamma_{n-1}, & n \text{ not prime,} \end{cases} \qquad (n \geq 5). \tag{5}$$

Theorem 3.4 (nearest internal positions, with an explicit gap)**.** *For every $n \geq 4$,*

$$N_{n,2} + \gamma_n \leq \varphi(L_n), \qquad \gamma_n > 0,$$

and consequently $N_{n,2} = N_{n,n-1} < \varphi(L_n)$, which is [1, Theorem 6.4]*. In closed form*

$$\gamma_n = \frac{L_n}{6} \prod_{5 \leq p \leq n,\ p \text{ prime}} \Bigl(1 - \frac{2}{p}\Bigr). \tag{6}$$

Proof sketch. Induction on n from the base $N_{4,2} = 2$, $\varphi(L_4) = 4$, $\gamma_4 = 2$. At a prime level $p = n+1$, (4) at $j = 2$ reads $N_{p,2} = (p-2)N_{n,2} + N_{n,1}$, and $N_{n,1} = \varphi(L_n)$, $\varphi(L_p) = (p-1)\varphi(L_n)$; multiplying the inductive hypothesis by $p-2$ and adding $\varphi(L_n)$ gives the claim with $\gamma_p = (p-2)\gamma_n$. At a level $n+1$ that is not prime, multiply the inductive hypothesis by ρ_{n+1} and use both halves of Proposition 3.3. Positivity of γ_n is immediate from (5), since $\rho_n \geq 1$ and $n-2 \geq 3$ at prime levels $n \geq 5$. The closed form (6) is the same induction: at a prime level p the period gains the factor p and the product gains $1 - 2/p$, whose product is $p-2$; at a level that is not prime the period gains ρ_n and the product is unchanged, because no new prime enters (Proposition 2.4(ii)). □

Remark 3.5. The formal development proves the recursive form (5); the closed form (6) is the elementary induction just given and was checked numerically against (5) for every $4 \leq n \leq 66$. The gap is of the true order of magnitude but loose by a constant factor: at $n = 66$ one has $\gamma_{66} = 0.174\,\varphi(L_{66})$ against the true deficiency $\varphi(L_{66}) - N_{66,2} = 0.569\,\varphi(L_{66})$, and $\gamma_n \leq \varphi(L_n) - N_{n,2}$ at every $4 \leq n \leq 66$. By a Mertens-type estimate, (6) gives $\gamma_n/L_n \asymp 1/\log^2 n$ and hence $\gamma_n/\varphi(L_n) \asymp 1/\log n$, whereas the computed ratio $(\varphi(L_n) - N_{n,2})/\varphi(L_n)$ shows no such decay: it lies between 0.549 and 0.580 at every level $25 \leq n \leq 66$.

Corollary 3.6 (the least counterexample; [1, Corollary 6.5])**.** *Let $m \geq 4$, suppose endpoint dominance holds at level m and fails at level $m+1$. Then $m+1$ is not prime, and there is a j with $3 \leq j \leq m-1$ and $N_{m+1,j} \geq \varphi(L_{m+1})$.*

Proof. Primality of $m+1$ is excluded by Proposition 3.2. A failure occurs at some internal j, and $j \in \{2, m\}$ is excluded by Theorem 3.4 together with the palindrome of Proposition 2.3. ☐

4. Exact fiber-size vectors to $n = 66$, and the frontier

The computation. Direct enumeration of $[0, L_n)$ stops being possible around $n = 22$ ($L_{22} = 232\,792\,560$). To go further we use the structure that makes the fiber sizes well defined in the first place. Fix the base $B = 18$, $L_B = L_{18} = 12\,252\,240$. For a prime $p \leq n$ put $A_p = \max\{a : p^a \leq n\}$ and $B_p = \max\{b : p^b \leq B\}$. The Chinese remainder theorem together with the p-adic digit expansion identifies

$$q \bmod L_n \;\longleftrightarrow\; \big(r,\ (d_p)_{A_p > B_p}\big), \qquad r = q \bmod L_B, \qquad q \bmod p^{A_p} = (r \bmod p^{B_p}) + p^{B_p} d_p,$$

with $d_p \in [0, p^{A_p - B_p})$. Level k consults a digit exactly when $p^{B_p+1} \mid k$; for $n \leq 66$ at most one digit is consulted at any single level (two would require a product of two such primes, at least $19 \cdot 23 = 437 > 66$), and at most seven digits are simultaneously live. The program enumerates r over $[0, L_B)$ and carries, for each r, a set of *blocks*: a surviving digit set per live prime (a 64-bit mask), the current state F, and a weight. A level not consulting a digit advances every block deterministically; a level consulting the digit of p splits each block's p-set at the threshold F (the threshold masks are built once per pair (r, k) by a counting sort over $q \bmod k$, so a split is two bitwise ANDs); when the last multiple of p has passed, that coordinate is summed out and blocks merge. Block counts stay below about 100. Every quantity is an exact integer. This is a general-n implementation of the idea used, hand-specialised to $n = 49$, by the program of [2]; it was written after reading that program and is not a port of it.

Cross-checks. Three independent recomputations of published numbers were run before any new level was believed. A direct enumeration over $[0, L_{22})$ reproduces the block program's output for $1 \leq n \leq 22$; the block program reproduces the complete published table [2] for $1 \leq n \leq 49$, every digit of all 49 rows, the last of which is 49 integers of up to 21 digits; and the proof assistant recomputes the rows $3 \leq n \leq 14$ from (1) itself (Proposition 2.5). Levels $50 \leq n \leq 64$ were also recomputed by a separate run with a different thread count and target level. In addition, the program checks at every level the three invariants that Propositions 2.4(iv), 2.2 and 2.3 predict: the vector sums to L_n, both endpoint entries equal $\varphi(L_n)$, and the vector is a palindrome. All hold at every level up to 66.

Theorem 4.1 (the frontier)**.** *Endpoint dominance holds at every level $4 \leq n \leq 67$. Consequently, if Problem 1.1 has a counterexample and n is the least level at which it fails, then*

$$n \geq 68, \qquad n \textit{ composite}, \qquad 3 \leq j \leq n-2.$$

Proof. For $4 \leq n \leq 66$ the statement is the exhaustive computation described above: the exact vector $(N_{n,1}, \ldots, N_{n,n})$ was computed at each level and compared entry by entry with $\varphi(L_n)$. The search is complete in the sense that the block decomposition partitions all of $\mathbb{Z}/L_n\mathbb{Z}$ — at $n = 66$ that is $L_{66} = 1\,182\,266\,884\,102\,822\,267\,511\,361\,600$ classes, organised as L_{18} outer iterations carrying at most about a hundred weighted blocks each — and every count is an exact integer. Level 67 is prime, so Proposition 3.2 applied at $m = 66$ gives endpoint dominance at level 67 from the computed level 66, with no computation. The description of the least counterexample is Corollary 3.6. ☐

Remark 4.2 (what is machine-checked here and what is computed)*.* The boundary matters and we state it plainly. Proposition 3.2, and hence the step from level 66 to level 67, is a formally verified theorem valid for every n (Section 7). The fiber-size vectors for $15 \leq n \leq 66$ are *not* kernel-checked: they are the output of a program in C, trusted on the strength of the three agreements listed above — with a direct enumeration to $n = 22$, with the published table to $n = 49$, and with the proof assistant's own recomputation to $n = 14$. Formally verifying the block algorithm would require the digit bijection $\mathbb{Z}/L_n\mathbb{Z} \cong \mathbb{Z}/L_{18}\mathbb{Z} \times \prod_p \mathbb{Z}/p^{A_p - B_p}\mathbb{Z}$ and a correctness proof of the splitting step; that is a formalization project of its own and was not attempted.

Remark 4.3 (cost, and why the computation stops at 66). Measured single-machine cost of this implementation: 10.7 CPU-seconds to reach $n = 22$, 243 to reach $n = 49$, 307 at $n = 50$, 424 at $n = 56$, 4 327 at $n = 64$ and about 5 700 at $n = 66$. The tenfold jump between 56 and 64 is caused by the prime 19 acquiring a third multiple at level 57 and by seven digits being live at once; extrapolating, $n = 72$ costs on the order of 10 to 40 CPU-hours. A second, harder limit sits exactly at $n = 67$: the digit set of a prime $p > B$ is held in a 64-bit mask, so $p = 67$, which needs 67 digit values, does not fit and the program refuses to run. It is precisely at that level that primality makes the computation unnecessary. Both limits are fixable — two-word masks, or the larger base L_{19} — and neither is cheap; we have not done it.

The table below records, at each new level, whether the level is prime, the number of decimal digits of L_n, and the ratio $N_{n,2}/\varphi(L_n)$ — by Theorem 5.1(ii) the largest internal fiber relative to the endpoint value.

n	type	digits of L_n	$N_{n,2}/\varphi(L_n)$	n	type	digits of L_n	$N_{n,2}/\varphi(L_n)$
50	composite	22	0.4340	59	prime	25	0.4365
51	composite	22	0.4321	60	composite	25	0.4330
52	composite	22	0.4290	61	prime	27	0.4424
53	prime	24	0.4400	62	composite	27	0.4391
54	composite	24	0.4355	63	composite	27	0.4377
55	composite	24	0.4339	64	composite	28	0.4356
56	composite	24	0.4313	65	composite	28	0.4343
57	composite	24	0.4298	66	composite	28	0.4312
58	composite	24	0.4266				

At the top level the three quantities the problem compares are

$$\begin{aligned} L_{66} &= 1\,182\,266\,884\,102\,822\,267\,511\,361\,600, \\ \varphi(L_{66}) &= 155\,571\,368\,459\,360\,639\,385\,600\,000, \\ N_{66,2} &= 67\,085\,480\,541\,822\,273\,831\,658\,050. \end{aligned}$$

The ratio $N_{n,2}/\varphi(L_n)$ is flat over the whole computed range — 0.4382 at $n = 49$, 0.4340 at $n = 50$, 0.4312 at $n = 66$ — so nothing in the data suggests that the largest internal fiber is approaching the endpoint value.

5. The position of the internal maximum

The computation of Section 4 produces, at each level, not only a verdict but the whole vector, and the whole vector says something the verdict does not.

Theorem 5.1 (the internal maximum, and what it implies)**.** (i) *For every n with $4 \le n$: if $N_{n,j} \le N_{n,2}$ for every j with $1 < j < n$, then*

$$N_{n,j} + \gamma_n \le \varphi(L_n) \qquad (1 < j < n),$$

and in particular endpoint dominance holds at level n.

(ii) *The hypothesis of (i) holds at every level $3 \le n \le 66$: the internal maximum* $\max_{1<j<n} N_{n,j}$ *is attained at $j = 2$, and hence also at $j = n - 1$.*

Consequently, on the whole range where the fiber-size vector has been computed, endpoint dominance is a corollary of a statement about the position *of one maximum, and Problem 1.1 is equivalent to that positional statement holding for all n.*

Proof. (i) is Theorem 3.4 and the hypothesis: $N_{n,j} \le N_{n,2} \le \varphi(L_n) - \gamma_n$ for internal j, and $\gamma_n > 0$. This part is proved for every n and is machine-checked.

(ii) rests on exhaustive computation, on two different ranges. For $4 \le n \le 14$ the Boolean $\left[\forall j,\ 1 < j < n \Rightarrow N_{n,j} \le N_{n,2}\right]$ was evaluated inside the proof assistant from the recurrence (1), by enumeration of the full period at each of the eleven levels. For $15 \le n \le 66$ it is read off the exact vectors of Section 4, with the same trust boundary as Theorem 4.1 (Remark 4.2); the level $n = 3$ is trivial, having a single internal position. In all 64 rows with $3 \le n \le 66$ — the 47 of them lying in the published range [2] and the 17 new ones — the internal maximum sits at $j = 2$. □

Remark 5.2. Part (ii) is a statement about the source's own data as much as about ours: the program of [2] prints the position of the internal maximum in every row (its `argmax` field), and its checker recomputes that position and rejects the row if it disagrees, but the text of [1] never records where the maximum sits. Its assertions about maxima are the statement of Problem 1.1 and its restatements in that paper's introduction and conclusion, together with the appendix remark that the checker verifies "the reported internal maximum"; the value $j = 2$ is named nowhere. What part (i) adds is that this omitted regularity is not a curiosity but the whole problem: no further arithmetic input is needed to pass from it to endpoint dominance, and the passage even carries a quantitative margin. We emphasise what is *not* claimed: the position of the internal maximum is not proved for general n here, and part (i) makes the positional statement equivalent in strength to Problem 1.1, not easier than it.

Remark 5.3 (the next position). The natural next target is a single interior position: proving $N_{n,3} < \varphi(L_n)$ for all n would shrink the frontier of Theorem 4.1 to $4 \le j \le n-3$. The induction of Theorem 3.4 does not transfer, because the leakage term $T_{n,2} = \#\{q \bmod L_n : F_{n-1}(q) = 2,\ q \bmod n < 2\}$ — a transition count in the sense of [1, Proposition 6.6] — must be controlled, and the crude bound $T_{n,2} \le 2L_n/n$ is not enough: relative to L_n it sums over levels like $2\log n$, against a gap γ_n/L_n that decays like $1/\log^2 n$ (Remark 3.5). Structure is available: the condition $F_{n-1}(q) = 2$ forces the single increment to occur at some prime level p, and then $q \equiv 0 \pmod n$ forces $n = p^2$, while $q \equiv 1 \pmod n$ forces every proper divisor > 1 of n to be at most p. We have not turned this into a bound.

6. The majorization defect

For composite $n \ge 4$ set, following [1, Section 6],

$$x^{(n)} = (\rho_n N_{n-1,1}, \dots, \rho_n N_{n-1,n-1}, 0), \qquad y^{(n)} = (N_{n,1}, \dots, N_{n,n}),$$

$$D_{n,k} = \sum_{i=1}^{k} \big(x^{(n)\downarrow}\big)_i - \sum_{i=1}^{k} \big(y^{(n)\downarrow}\big)_i \qquad (1 \le k \le n),$$

where $\downarrow$ denotes decreasing rearrangement. Both vectors sum to L_n, so $D_{n,n} = 0$. The source reports [1, Proposition 6.9] that $D_{n,k} = 0$ for $k \in \{1, 2, n\}$ and $D_{n,k} > 0$ for $3 \le k < n$, at every composite $4 \le n \le 49$, and observes that $D_{n,3} \ge 0$ at every composite level would settle Problem 1.1: if an internal coordinate of $y^{(n)}$ reached $\varphi(L_n)$, the three largest coordinates of $y^{(n)}$ would sum to at least $3\varphi(L_n)$, while $x^{(n)}$ has only two coordinates that large.

Theorem 6.1 (the defect past the published range)**.** (i) *At every composite level* $4 \le n \le 66$*:* $D_{n,k} = 0$ *for* $k \in \{1, 2, n\}$ *and* $D_{n,k} > 0$ *for* $3 \le k < n$*.*

(ii) *At every composite level* $4 \le n \le 66$*,*

$$D_{n,3} = \rho_n N_{n-1,2} - N_{n,2},$$

which is exactly the slack in the composite lift of Proposition 3.3.

(iii) *Over the extended range the smallest ratio is attained at* $(n, k) = (65, 3)$*:*

$$\min_{\substack{4 \le n \le 66,\ n \text{ composite} \\ 3 \le k < n}} \frac{D_{n,k}}{L_n} = \frac{D_{65,3}}{L_{65}} = \frac{5\,854\,222\,475\,316\,720}{32\,853\,888\,335\,968\,339\,211} \approx 1.7818963818 \times 10^{-4},$$

with $D_{65,3} = 210\,667\,708\,307\,757\,359\,232\,000$ *and* $L_{65} = 1\,182\,266\,884\,102\,822\,267\,511\,361\,600$*. The corresponding minimum over the published range* $4 \le n \le 49$ *is*

$$\frac{D_{39,3}}{L_{39}} = \frac{2\,043\,072\,801\,000}{5\,342\,931\,457\,063\,200} = \frac{289\,035}{755\,868\,412} \approx 3.823879863 \times 10^{-4},$$

reproduced here exactly from the recomputed vectors; the record is therefore beaten, sixteen levels past the published range, by a factor 2.146.

Proof. All three parts rest on exhaustive computation over a finite range. For composite $n \le 14$ the whole of (i) and (ii) — the three vanishing partial sums, the $n-3$ strict positivities, and the identity of (ii) — was evaluated inside the proof assistant directly from the definitions, the fiber sizes themselves being recomputed from (1); the set of levels tested there is $\{4, 6, 8, 9, 10, 12, 14\}$, which was itself checked to be the list of composite n in range, so the check is not vacuous. For $15 \le n \le 66$, (i)–(iii) are computed from the exact vectors of Section 4 in exact rational arithmetic, with the trust boundary of Remark 4.2. The search in (iii) is over all pairs (n, k) with $n \le 66$ composite and $3 \le k < n$, which is 1 568 pairs; the published value at $(39, 3)$ was recomputed as a control before the extended range was examined. □

Remark 6.2 (what (ii) does to the proposed route). Part (ii) is not an accident of the range. Suppose that at levels $n-1$ and n (n composite) endpoint dominance holds and the internal maxima sit at $j = 2$. Then the two largest coordinates of $x^{(n)}$ are $\rho_n \varphi(L_{n-1}) = \varphi(L_n)$, and its third largest is $\rho_n N_{n-1,2}$; the two largest coordinates of $y^{(n)}$ are $\varphi(L_n)$ and its third largest is $N_{n,2}$. Subtracting gives $D_{n,3} = \rho_n N_{n-1,2} - N_{n,2}$, which is nonnegative by Proposition 3.3. In other words, on any range where the internal maxima are at $j = 2$, the inequality $D_{n,3} \ge 0$ that [1] proposes to prove is already a theorem — and conversely, the route through $D_{n,3}$ does not avoid the positional question of Section 5, it re-encodes it. The general identification is not formalized here; only its instances at composite $n \le 14$ are.

7. Verification

Every statement of this paper that is valid for general n — Propositions 2.1, 2.2, 2.3, 2.4, 3.1, 3.2, 3.3, Theorem 3.4, Corollary 3.6, Theorem 5.1(i), and the step from level 66 to level 67 inside Theorem 4.1 — has been formally verified in Lean 4 (toolchain `v4.32.0`) against *Mathlib* [6, 7]; the development contains no `sorry`. Of its 66 theorems, 47 — among them every statement just listed, together with the recursive definition and the positivity of γ_n — depend on no axioms beyond `propext`, `Classical.choice` and `Quot.sound`, so they are checked by the kernel alone; the base case $n = 4$ of the induction of Theorem 3.4 is closed by kernel evaluation, so that whole chain is kernel-checked. The remaining 19 theorems are exactly the finite tables: the twelve vectors and the dominance check over $4 \le n \le 14$ of Proposition 2.5, the internal-maximum check over $4 \le n \le 14$ (Theorem 5.1(ii)), the defect check over composite $4 \le n \le 14$ (Theorem 6.1(i),(ii)) and the four controls of Proposition 2.6; each delegates one Boolean evaluation to compiled code through Lean's `native_decide`, and each is an enumeration of one full period per level, from (1) itself. What is *not* formally verified is stated in Remark 4.2: the exact fiber-size vectors for $15 \le n \le 66$, on which Theorems 4.1, 5.1(ii) and 6.1 depend for those levels, are the output of a program in C, checked against the three independent recomputations listed there. Repository reference to be added at submission.

References

[1] L. Chen, *Exact Periodicity, Surjectivity, and a Haar Limit Law for a Restarting Josephus Process*, preprint arXiv:2608.09092v1 [math.CO], 10 August 2026. Numbers cited here are those of v1, the only version as of 2026-08-28.

[2] L. Chen, ancillary files of arXiv:2608.09092v1, in the directory `anc/`: the generator `tools/exact_counts_49.cpp`, the checker `tools/check_majorization.py`, and the exact output `supplement/results/exact_counts_49.stdout`, which is the complete fiber-size table for $1 \le n \le 49$.

[3] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics*, 2nd edition, Addison-Wesley, 1994, Section 1.3.

[4] L. Halbeisen and N. Hungerbühler, *The Josephus problem*, J. Théor. Nombres Bordeaux **9** (1997), no. 2, 303–318.

[5] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, `https://oeis.org`, entry `A128982`.

[6] L. de Moura and S. Ullrich, *The Lean 4 Theorem Prover and Programming Language*, in: Automated Deduction — CADE-28, Lecture Notes in Computer Science 12699, Springer, 2021, pp. 625–635.

[7] The mathlib Community, *The Lean Mathematical Library*, in: Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381.