Aicumene:2026.053v1 [econ.TH] November 2026

Multi-Agent Verification Games III: Verification Networks, Loops, and the Endogenous Price of Independence

Aicumene Preprint

Alexandra Bernadotte
Aicumene*

Ivan Menshikov
Aicumene

November 2026 — working paper, Part III (SSRN preprint)

Abstract

In the first two parts of this series the premium d that an agent pays for a source-independent verification was a parameter, and the crowding-out wedge $pcL < d < npcL$ was a condition on that parameter. Here the premium is endogenous: agents sit in a consultation network, a shared verification is one question to a neighbour, and an independent one means reaching the origin of the record along the chain of delegations, which may be long, may require finding a sink under local information, and may not exist when the chain closes into a loop. Three results follow. First, in the delegation model of Part I, in which each agent goes to the origin with probability θ and otherwise asks a random colleague, the premium of an agent is her depth times the price of a question, the depth is asymptotically geometric, and the wedge of Part I becomes the band $\kappa/(npcL+\kappa) < \theta < \kappa/(pcL+\kappa)$ on the share of agents who go to the origin. Second, when agents know only their own neighbourhood, the price of the origin on the h-cube is the query complexity of finding the sink of a unique-sink orientation: at least of order $h^2/\log h$ and at most $O(1.61^h)$, so that in large organizations even the social optimum verifies through copies; and of the 744 unique-sink orientations of the 3-cube, in which every sub-community has one responsible agent, 16 contain a loop. Third, a network procurement game in which each agent decides whether to go to the origin before the random consultation structure is drawn has, for $d < pcL$, a unique pure equilibrium count of origin-checkers $k^* = a^*\sqrt{n}\,(1+o(1))$, with a^* the solution of $\Phi(a) = d/(pcL)$ for the explicit scaling function $\Phi(a) = \int_0^\infty u\, e^{-au-u^2/2}\, du$; in it a share $d/(pcL)$ of the agents verify through self-confirming loops, every agent pays the premium in expected loss without obtaining independence, and the architecture contains loops with probability tending to one. The social optimum sends $(2n^2pcL/d)^{1/3}$ agents to the origin, a share of agents that still vanishes, and the welfare loss per agent tends to d. For $d \geq pcL$ nobody goes to the origin while the optimum sends someone whenever $d < pcL\sqrt{\pi n/2}$, the audience of one origin-checker. When the architecture is known, the pure equilibria are exactly the minimum feedback vertex sets of the delegation map, one agent per loop, of expected size $\frac{1}{2}\log n + O(1)$: the loops persist in the unknown case not because independence is too expensive but because nobody knows she is on a loop, and the cheapest mandate is informational. A protocol for the same game with language-model agents is given. The results are conditional on the stated model, and every finite claim is checked by exact computation and enumeration.

Keywords: multi-agent systems; verification networks; evidentiary independence; costly verification; random mappings; unique-sink orientations; feedback vertex set; public goods; information economics; artificial intelligence.
JEL: D82, D83, D85, H41.

*Aicumene: `https://aicumene.com/`; this paper: `https://research.aicumene.com/papers/game-theory-3/`. Corresponding author: bernadotte@aicumene.com.
© 2026 Alexandra Bernadotte and Ivan Menshikov. Licensed under CC BY-NC-ND 4.0.

Aicumene:2026.053v1 [econ.TH] November 2026

1 Introduction

Part I of this series [1] established a floor on what verifiers who share a source can detect, $FA \geq pc(1 - FR/(1-p))$, and a wedge of prices, $pcL < d < npcL$, on which every pure equilibrium of a procurement game buys a shared-source check although one independent check would be cheaper for the group; Part II [20] showed the same wedge in sequential reliance, in verification markets and across provenance structures. In both, the independence premium d was a number given to the agents. This paper asks where the number comes from.

The answer proposed here is topological. Verifiers live in a network of consultation: an agent checks a claim by asking a neighbour, who may have asked a neighbour in turn, and the chain of delegations either ends at an agent who went to the origin of the record, the vault of the epigraph of Part I, or it does not end at all. A shared verification is one question. An independent one is reaching the origin, and its price depends on three things that Parts I and II held fixed: how far the origin is along the chain (Section 3), whether the agent can find it at all when she sees only her own neighbourhood (Section 4), and whether her chain closes into a loop, in which case the origin is unreachable and the agents on the loop confirm one another's records forever (Sections 5 and 6). Proposition 2 of Part I counted the architectures without loops and found them rare; here the loops are given their economics.

Contributions. (i) *The premium is the depth.* In the delegation model of Part I the independence premium of an agent at depth t is $\kappa(t-1)$, the depth is asymptotically geometric with parameter θ, the share of agents who go to the origin, and the wedge of Part I becomes a band of depths and a band $\kappa/(npcL+\kappa) < \theta < \kappa/(pcL+\kappa)$ of shares (Proposition 1). (ii) *The price of the origin under local information.* On the h-cube with a unique responsible agent in every sub-community, finding the origin is finding the sink of a unique-sink orientation with a vertex oracle; the known bounds, $\Omega(h^2/\log h)$ queries for every deterministic rule and $O(1.61^h)$ for the best known one, translate into a premium that outgrows $npcL$ (Theorem 1), and of the 744 such architectures of the 3-cube 16 contain a loop (Proposition 2). (iii) *Loops in time.* The entrenchment theorem of Part II holds on a loop with c replaced by the share $c(Z)$ of errors that originate on it, and on a closed loop the shared verdict carries no information at all (Proposition 3). (iv) *The network procurement game.* The main result, Theorem 2, characterizes the pure equilibria and the social optimum of the game in which each agent decides whether to pay the premium before the random consultation structure is drawn: for $d < pcL$ the equilibrium count of origin-checkers is $a^*\sqrt{n}$, the trapped share is $d/(pcL)$, the optimum count is $(2n^2pcL/d)^{1/3}$ and the welfare loss per agent tends to d; for $d \geq pcL$ nobody goes while the optimum sends someone whenever $d < pcL\sqrt{\pi n/2}$ (Corollary 1). The proofs rest on an exact formula for the probability that a random delegation chain never reaches a sink (Lemma 1) and its scaling limit Φ. (v) *Mandates.* With a known architecture the equilibria are the minimum feedback vertex sets of the delegation map, one agent per loop; the expected number of loops is $\frac{1}{2}\log n + O(1)$, so the cheapest intervention is to tell the agents on loops where they are (Theorem 3, Proposition 4). (vi) A protocol for the same game with language-model agents, to be run under the criterion of [15] (Section 8).

Road map. Section 2 fixes the network, the delegation model and the error model, and recalls the loop lemma of Part I. Sections 3–5 give results (i)–(iii). Section 6 states and proves Theorem 2, Section 7 the results on known architectures and mandates. Section 8 gives the experimental protocol, Section 9 the related work, Section 10 the open problems, and Section 11 concludes. All numbers in the text are produced by the scripts of Section 11.

2 The network

Agents, records, errors. There are n agents, each holding a copy of one record, and an origin O, the vault, holding the true record. A claim is wrong with probability p; conditional on W, with probability c the error is in the copies and not in the vault, the shared-source event E of Part I, and with probability $1-c$ it is idiosyncratic to the claim and visible to any check. The losses are L for accepting a wrong claim and ℓ for rejecting a correct one, with $pL < (1-p)\ell$ (accept by default) as in Parts I and II; own errors of verifiers are set to $q = 0$ throughout, the robustness of Section 6 of Part I applying verbatim.

Aicumene:2026.053v1 [econ.TH] November 2026

Table 1: Notation.

Symbol	Meaning
p, c, L, ℓ	probability of a wrong claim; share of errors shared by the copies; losses (Parts I–II)
κ, d	price of one question to a neighbour; independence premium of a direct visit to the origin
$\delta = d/(pcL)$	the premium in units of the value of independence to one agent
G, D, O	consultation graph; architecture (verification digraph); the origin
k, $\theta = k/n$	number and share of agents who go to the origin (the sinks of D)
t, D_i	depth: number of arcs from an agent to a sink
$F(k, n)$	probability that a random delegation chain never reaches a sink (Lemma 1)
$\Phi(a)$, a^*	scaling limit of $F(a\sqrt{n}, n)$; the solution of $\Phi(a^*) = \delta$
k^*, k_{opt}	equilibrium and socially optimal numbers of origin-checkers (Theorem 2)
$\nu(D)$	number of directed cycles of D, the size of the minimal mandate (Proposition 4)
Q_h	the h-cube; a unique-sink orientation (USO) has one sink in every face
Λ_Z	entrenchment threshold on a loop, $(1-p)\ell/(p\,c(Z)L)$ (Proposition 3)

Consultation and delegation. A *consultation graph* G on the agents says who may ask whom. An *architecture* is a verification digraph D: an arc $i \to j$ means that i verifies by asking j, and an agent with no outgoing arc verifies at the origin and is a *sink*. The *delegation model* of Proposition 2(iii) of Part I is the case of the complete consultation graph in which every agent either goes to the origin or asks exactly one other agent; the *random delegation architecture* with k sinks has the k sinks fixed and every other agent asking a uniformly random other agent, independently. The *chain* of agent i is $i \to j \to \cdots$; it either reaches a sink, in which case we say that i *reaches the origin*, at *depth* equal to the number of arcs, or it repeats an agent and then runs on a directed cycle forever, in which case i is *trapped*. Agents on the cycle and agents whose chain enters it are trapped together; a cycle together with the agents that lead into it is a *loop* and the trapped agents form its *basin*.

What a check is worth. A verification obtained through a chain that reaches the origin detects every error, because the sink compares the claim with the true record and the verdict is passed back along the chain; its decision loss is 0. A verification obtained on a loop compares the claim with records that carry the same error: by Lemma 2 of Part I (error propagation on a loop), the verdicts on a cycle Z have, on the event that the error originates on Z, the same joint law as on a correct claim, and when every copying error of the claims handled on Z is Z's own the floor of Part I applies with $c_{\mathrm{src}}(Z) = 1$ for those errors. In the notation of Part I a trapped agent's check is a shared check, with decision loss pcL, and a check that reaches the origin is an independent one, with decision loss 0; an agent who does not check at all has loss pL. Prices: asking a neighbour costs κ, going to the origin directly costs $\kappa + d$ with $d > 0$ the independence premium, and walking the chain costs κ per arc. Throughout, $\kappa < pL(1-c)$, so that a check is worth buying even when it is shared.

Table 1 collects the notation.

3 The premium is the depth

Suppose first that agents know the architecture and can walk their chain: an independent verification for agent i at depth t means t questions, the last one answered at the origin, so its price is κt and the independence premium, relative to one shared question, is $d_i = \kappa(t-1)$. The wedge of Part I, $pcL < d < npcL$ for a verdict relied on by n receivers, is then a condition on where the agent sits.

Proposition 1 (Depth, the geometric law, and the band in θ)**.** *(i) In an architecture in which agent i reaches the origin at depth t, the independence premium of walking the chain is $d_i = \kappa(t-1)$, and i is crowded out in the sense of Theorem 3 of Part I iff $pcL/\kappa < t - 1 < npcL/\kappa$. (ii) In the random delegation architecture with $k = \theta n$ sinks, $0 < \theta < 1$ fixed, the depth of a non-sink agent who reaches the origin satisfies*

$$\Pr(D_i = t) \;=\; \frac{k}{n-1}\prod_{s=1}^{t-1}\Big(1 - \frac{k+s}{n-1}\Big) \;\longrightarrow\; \theta(1-\theta)^{t-1} \qquad (n \to \infty),$$

Aicumene:2026.053v1 [econ.TH] November 2026

the probability of being trapped tends to 0*, and the expected depth tends to* $1/\theta$*. (iii) Hence the expected premium tends to* $\kappa(1-\theta)/\theta$*, and the typical agent is crowded out iff*

$$\frac{\kappa}{npcL+\kappa} \;<\; \theta \;<\; \frac{\kappa}{pcL+\kappa}. \tag{1}$$

Proof. (i) is the definition of the premium inserted in Theorem 3 of Part I. (ii) Follow the chain from i. When it stands at its t-th distinct non-sink agent, that agent's arc is uniform over the other $n-1$ agents and independent of the arcs seen so far: it points to a sink with probability $k/(n-1)$, to one of the $t-1$ agents already visited with probability $(t-1)/(n-1)$, and to a fresh non-sink agent otherwise; the displayed product is the probability of $t-1$ fresh steps followed by a sink. With $k=\theta n$ the factors $1-(k+s)/(n-1)$ converge to $1-\theta$ uniformly for $s \le t$, giving the geometric limit, and the trapped probability is $\sum_t \frac{t-1}{n-1}\prod_{s<t}(1-\frac{k+s}{n-1}) \le \sum_t \frac{t-1}{n-1}(1-\theta)^{t-1} = \frac{1-\theta}{(n-1)\theta^2} \to 0$. (iii) follows from (i), (ii) and $\kappa(1/\theta-1) \in (pcL, npcL)$. □

Figure 1 shows the exact depth law against the geometric one and the share of agents in the wedge as a function of θ. For the running example of Part I ($p=0.1$, $c=0.5$, $L=100$, $\kappa=3$, so $pcL=5$), the typical agent is crowded out for $0.167<\theta<0.375$ when three receivers rely on each verdict; the share of all agents in the wedge is 0.22 at $\theta=0.1$ and 0.22 at $\theta=0.5$ and peaks between. The band (1) is the wedge of Part I read backwards: it says how many agents must go to the origin for the premium of the others to fall below the private value of independence. Nothing in it yet says who will.

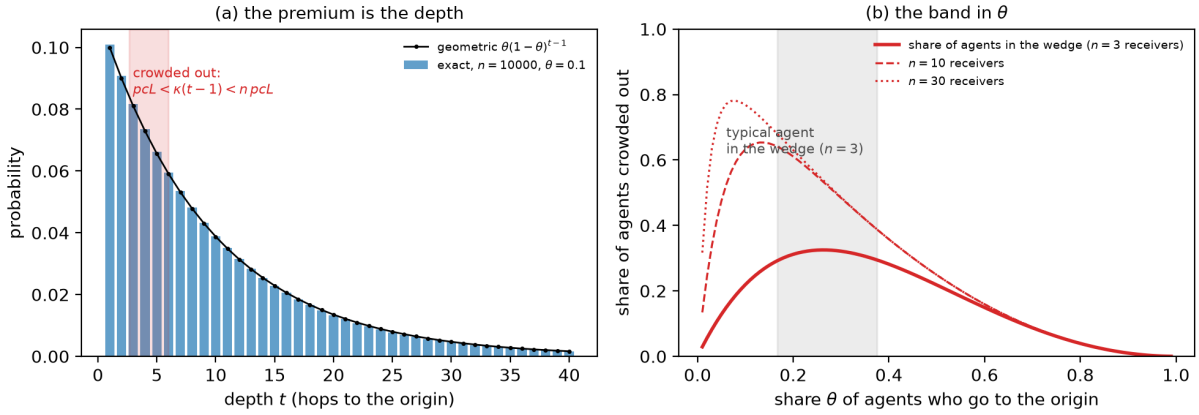

Figure 1: The premium as depth. (a) Exact depth distribution of a non-sink agent in the random delegation architecture ($n=10^4$, $\theta=0.1$) against the geometric law; the shaded depths are those at which the agent is crowded out in the running example with three receivers per verdict. (b) Share of agents in the wedge against θ for 3, 10 and 30 receivers; shaded, the band (1) in which the typical agent is crowded out for 3 receivers. Generated by `depth_band.py` and `make_figures.py`.

4 Local information: the price of the origin on the cube

Proposition 1 assumed that the agent can walk her chain. Under local information she cannot: she sees the orientation of the arcs at her own position and learns the rest only by asking. Part I used the h-cube Q_h as the consultation graph of agents indexed by h binary attributes, and the *unique-sink orientations* (USO) of [11], in which every face has exactly one sink, as the architectures in which every sub-community has one responsible agent. Reaching the origin is then the problem of finding the sink of a USO with a *vertex oracle* that, asked about a vertex, returns the orientation of its h incident edges [24].

Theorem 1 (The price of the origin under local information)**.** *Let the architecture be a unique-sink orientation of* Q_h *and let an agent learn the orientation only through vertex queries at price* κ *each. Write* $t(h)$ *for the*

Aicumene:2026.053v1 [econ.TH] November 2026

number of queries that the best deterministic rule needs in the worst case over orientations and starting vertices. Then

$$\Omega\big(h^2/\log h\big) \;\le\; t(h) \;\le\; O\big(1.61^h\big),$$

the lower bound by [25] and the upper bound by [24]; on acyclic orientations the randomized Random-Facet rule needs $e^{O(\sqrt{h})}$ expected queries [26] while Random-Edge, the rule "ask any colleague who points away from you", can need exponentially many [27]. Consequently, for fixed κ, p, c, L and n there is h_0 such that for every $h \ge h_0$ there are architectures of Q_h and starting positions at which the independence premium $\kappa(t-1)$ of the best deterministic rule exceeds $npcL$: on them the social optimum of the procurement game of Part I verifies through copies, and the wedge is empty not because independence is cheap but because it is unreachable at any admissible price.

Proof. The bounds are the cited theorems: [24] give an algorithm that finds the sink of any USO of Q_h with $O(1.61^h)$ vertex evaluations, [25] prove that every deterministic algorithm needs $\Omega(h^2/\log h)$ evaluations on some USO, [26] bound Random-Facet on acyclic USOs by $e^{O(\sqrt{h})}$ and [27] exhibit acyclic USOs on which Random-Edge needs $e^{\Omega(h^{1/3})}$ steps. For the consequence, $\kappa(t(h)-1) \ge \kappa(Ch^2/\log h - 1) > npcL$ for h large, and on an orientation and a start that attain the worst case the premium of the agent exceeds $npcL$, at which point one independent check costs the group more than the shared alternative $\kappa + npcL$ in Proposition 2 of Part I. □

Two remarks. The lower bound is a worst case over orientations; on a bottom-up architecture the agent who simply follows outgoing arcs reaches the origin, and the length of that walk is at most $2^h - 1$ and at least the graph distance to the sink, which is at most h: for $h = 3$ the worst downhill walk among the 728 acyclic USOs has 7 steps and the worst shortest walk 3 (script `uso_cycles.py`). The gap between h and $2^h - 1$ is the price of not knowing which colleague to ask. And unique responsibility does not protect against loops:

Proposition 2 (Loops under unique responsibility)**.** *All* 12 *unique-sink orientations of Q_2 are acyclic. Of the* 744 *unique-sink orientations of Q_3,* 728 *are acyclic and* 16 *contain a directed cycle, a share of* 0.022*; Figure 2(c) shows one. On a cyclic unique-sink orientation an agent who follows outgoing arcs can circle forever, and the floor of Lemma 2 of Part I applies to the agents on the cycle.*

Proof. Enumeration of all 2^{12} orientations of Q_3 and all 2^4 of Q_2, testing the unique-sink property on every face and acyclicity by depth-first search (`uso_cycles.py`). On Q_2 a directed cycle would be the whole 4-cycle, which has no sink. The last sentence is Lemma 2 of Part I. □

The share of cyclic unique-sink orientations for $h \to \infty$ is open (Section 10); both the USOs and the acyclic USOs number $h^{\Theta(2^h)}$ [11, 10], and the ratio is not known.

5 Loops in time: entrenchment on a cycle

Part II considered n receivers who decide in sequence, see the decisions of their predecessors and hold private independent signals, and showed that after a shared `ok` a cascade of acceptance starts at the first receiver whenever the likelihood ratio λ of a private signal is below $\Lambda = (1-p)\ell/(pcL)$. On a loop the same happens with c replaced by the loop's own share of the errors.

Proposition 3 (Entrenchment on a loop)**.** *Let the agents of a directed cycle Z of D decide in sequence on one claim under the premises of Lemma 2 of Part I, each seeing the public verdict obtained on Z, the decisions of her predecessors on Z and a private independent signal of likelihood ratio λ, and let $c(Z) = \Pr(\text{origin of the error} \in Z \mid W)$. Then Theorem 2 of Part II holds verbatim with c replaced by $c(Z)$ and Λ by $\Lambda_Z = (1-p)\ell/(p\,c(Z)L)$: for $\lambda < \Lambda_Z$ every agent of Z accepts regardless of her signal, $\Pr(W, A_{|Z|}) = p\,c(Z)$, and the $|Z|$ private signals go unused; for $\lambda > \Lambda_Z$ the outcome is decided by the first agent's signal alone. On a closed loop, one on which every copying error of the claims it handles originates, $c(Z) = 1$ and $\Lambda_Z = (1-p)\ell/(pL) = \Lambda_0$, the threshold a private signal needs when there is no verdict at all: the verdict obtained on a closed loop carries no information, and the loop entrenches exactly as an unchecked claim would.*

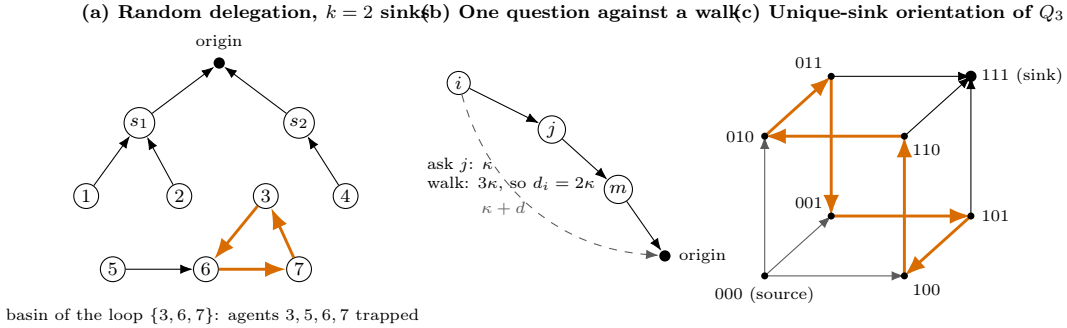

Figure 2: Verification networks; arrows point from an agent to the agent consulted. (a) A random delegation architecture with two sinks: agents 1, 2, 4 reach the origin at depth 2; agents 3, 6, 7 form a loop and agent 5 leads into it, so the basin of the loop has four trapped agents. (b) The premium as depth: one question to j costs κ and detects only idiosyncratic errors; walking to the origin costs 3κ, so the premium is 2κ, against the exogenous d of a direct visit. (c) The unique-sink orientation of the 3-cube of [11] with a directed 6-cycle (thick arrows), reproduced from Part I: every face has one responsible agent and the six agents of weight one and two verify only through one another; it is one of the 16 cyclic orientations among the 744 of Proposition 2.

Proof. By Lemma 2 of Part I the verdicts on Z are jointly indistinguishing on the event that the error originates on Z, whose conditional probability is $c(Z)$; this is the hypothesis of Lemma 1 of Part II with $c = c(Z)$, and the proof of Theorem 2 of Part II uses nothing else: the public odds after the shared ok are $p\,c(Z)/(1-p)$, a receiver rejects iff these odds times λ exceed ℓ/L, and the induction on the receiver index is unchanged. For $c(Z) = 1$ the odds after the verdict equal the prior odds $p/(1-p)$, which is the no-verdict case of the last sentence of Theorem 2 of Part II. □

Corollary 1 of Part II, the sequential crowding-out band $v < \kappa_I < nv$ with $v = pc\lambda L/(pc\lambda + 1 - p)$, transfers in the same way with $c(Z)$ in place of c and $|Z|$ in place of n: an independent check bought by the first agent of the loop would be worth $|Z|v$ to the loop and v to her. Section 6 turns this static band into the question of who, before the loops form, pays to be a sink.

6 The network procurement game

6.1 The game

Definition 1 (Network procurement game Γ_n^{net})**.** Players. *Agents $i = 1, \dots, n$, each holding a copy of one claim.* Stage 1 (simultaneous). *Each agent chooses O (go to the origin, price $\kappa + d$), A (ask, price κ) or $\varnothing$ (no check).* Stage 2 (nature). *Every agent who chose A is assigned a consultation partner drawn uniformly among the other $n-1$ agents, independently across agents; the resulting architecture D has the O-agents as sinks.* Verdicts. *An A-agent whose chain reaches a sink obtains the sink's verdict; an A-agent whose chain is trapped obtains a verdict that is shared on E in the sense of Section 2.* Stage 3. *Each agent takes the Bayes-optimal accept/reject decision on her own verdict.* Payoffs. $u_i = -price(a_i) - \mathcal{L}_i$, *with $\mathcal{L}_i = 0$ for an O-agent and for an A-agent who reaches the origin, $\mathcal{L}_i = pcL$ for a trapped A-agent and $\mathcal{L}_i = pL$ for $\varnothing$. Agents know n, the prices and the parameters, but not the realized partners; $\kappa < pL(1-c)$.*

The game differs from Γ_n of Part I in two ways. The verdict is not one public good consumed by all n receivers but a good that propagates along arcs: an origin-checker serves the agents whose chains lead to her, her *audience*, and nobody else. And the agents do not choose whom to ask; the consultation partner is drawn after the choices, which is the formal content of local information: an agent cannot direct her question to someone she knows to have reached the origin. Section 7 removes this assumption.

Aicumene:2026.053v1 [econ.TH] November 2026

Aicumene:2026.053v1 [econ.TH] November 2026

6.2 The probability of a loop

Lemma 1 (Trapping probability of a random delegation chain)**.** *In the random delegation architecture with k sinks among n agents, the probability that the chain of a given non-sink agent never reaches a sink is*

$$F(k,n) \;=\; \sum_{t=1}^{n-k} \frac{t-1}{n-1} \prod_{s=1}^{t-1} \Big(1 - \frac{k+s}{n-1}\Big), \tag{2}$$

with $F(0,n)=1$, $F(n,n)=0$ and F strictly decreasing in k. For $k \geq 1$,

$$F(k,n) \;\leq\; \frac{n-1-k}{k^2}. \tag{3}$$

If $k_n = \lfloor a\sqrt{n} \rfloor$ with $a > 0$ fixed, then

$$F(k_n,n) \;\longrightarrow\; \Phi(a) \;=\; \int_0^\infty u\, e^{-au-u^2/2}\, du \;=\; 1 - a\sqrt{\tfrac{\pi}{2}}\, e^{a^2/2}\, \mathrm{erfc}\big(a/\sqrt{2}\big), \tag{4}$$

a strictly decreasing function with $\Phi(0)=1$ and $\Phi(a) = a^{-2}(1+O(a^{-2}))$ as $a \to \infty$. If $k_n/\sqrt{n} \to \infty$ and $k_n/n \to 0$, then $F(k_n,n)\, k_n^2/n \to 1$. Finally the audience of a single origin-checker, the expected number of agents whose chain reaches her, is $(n-1)(1-F(1,n)) = \sqrt{\pi n/2}\,(1+o(1))$.

Proof. Formula (2) is the chain process of the proof of Proposition 1: the chain fails at its t-th distinct non-sink agent with probability $(t-1)/(n-1)$ after $t-1$ fresh steps, each fresh step having probability $1-(k+s)/(n-1)$ at the s-th agent. Monotonicity: every factor and the number of terms decrease in k. For (3) replace every factor by $1-k/(n-1) =: x$ and sum $\sum_{t\geq 1}(t-1)x^{t-1}/(n-1) = x/((n-1)(1-x)^2) = (n-1-k)/k^2$. For (4) write $t = u\sqrt{n}$ and use $\log(1-y) = -y + O(y^2)$: for $t \leq T_n = \sqrt{n}\log n$ the product equals $\exp\big(-(t-1)k_n/(n-1) - t(t-1)/(2(n-1)) + O(\log^3 n/\sqrt{n})\big) = e^{-au-u^2/2}(1+o(1))$ uniformly, so the partial sum up to T_n is a Riemann sum with mesh $n^{-1/2}$ of $\int_0^{\log n} u\, e^{-au-u^2/2} du$ and converges to $\Phi(a)$; the tail $t > T_n$ is at most $\sum_{t>T_n} \frac{t-1}{n-1}(1-\frac{k_n}{n-1})^{t-1}$, which is bounded by $\int_{\log n}^\infty u e^{-au(1+o(1))} du \to 0$. The closed form of the integral follows from $\frac{d}{du} e^{-au-u^2/2} = -(u+a)e^{-au-u^2/2}$, which gives $\Phi(a) = 1 - a\int_0^\infty e^{-au-u^2/2} du$, and the Gaussian integral is $e^{a^2/2}\sqrt{\pi/2}\,\mathrm{erfc}(a/\sqrt{2})$; the asymptotics are those of the Mills ratio. If $k_n/\sqrt{n} \to \infty$, the same expansion with $t = un/k_n$ gives $\prod_{s<t}(1-\frac{k_n+s}{n-1}) = e^{-u}(1+o(1))$ uniformly for $u \leq \log n$, because $t^2/n = u^2 n/k_n^2 \to 0$, so $F = \frac{n}{k_n^2}\int_0^\infty u e^{-u} du\,(1+o(1)) = \frac{n}{k_n^2}(1+o(1))$, the tail being handled as before. For the audience, $1-F(1,n) = \sum_t \frac{1}{n-1}\prod_{s<t}(1-\frac{1+s}{n-1})$ and the product is $e^{-u^2/2}(1+o(1))$ at $t = u\sqrt{n}$, so $(n-1)(1-F(1,n)) = \sqrt{n}\int_0^\infty e^{-u^2/2} du\,(1+o(1)) = \sqrt{\pi n/2}\,(1+o(1))$. □

The formula is exact and cheap, and it agrees with simulation: at $n = 1000$ and $k = 30$, $F = 0.3518$ against 0.3519 over 4000 sampled architectures (`network_game.py`); Figure 3(a) shows the convergence to Φ. The critical scale $\sqrt{n}$ is the birthday scale: a chain of length t has probability of order t^2/n of closing on itself and probability of order tk/n of meeting one of k sinks, and the two compete when $k \asymp \sqrt{n}$. With fewer sinks almost every chain closes before it finds one; with many more almost every chain finds one.

6.3 Equilibria, optimum, and the price paid for nothing

Write $\delta = d/(pcL)$ for the premium in units of the value of independence to one agent.

Theorem 2 (Equilibria and optimum of the network procurement game)**.** *In Γ_n^{net} with $\kappa < pL(1-c)$ and generic $d > 0$:*

1. (Pure equilibria.) *No agent chooses $\varnothing$. A profile with k agents choosing O is a pure Nash equilibrium iff $F(k,n) \leq \delta \leq F(k-1,n)$; since F is strictly decreasing, the equilibrium number of origin-checkers is unique, $k^* = \min\{k \geq 0 : F(k,n) \leq \delta\}$, and the equilibria are the $\binom{n}{k^*}$ assignments of the k^* roles. In particular $k^* = 0$, nobody goes to the origin and everybody is trapped, iff $\delta \geq 1$, that is $d \geq pcL$.*

Aicumene:2026.053v1 [econ.TH] November 2026

2. (Scaling.) *If $\delta < 1$, then $k^*/\sqrt{n} \to a^*$, the unique solution of $\Phi(a^*) = \delta$, the share of trapped agents $F(k^*, n) \to \delta$, and the expected cost of an A-agent, $\kappa + pcL\,F(k^*, n)$, tends to $\kappa + d$: in the limit every agent pays the premium in expected loss, and $1 - \delta$ of them obtain independence by asking while δ of them obtain nothing. The probability that the architecture is bottom-up tends to* 0.
3. (Optimum.) *The socially optimal number k_{opt} of origin-checkers minimizes $C(k) = n\kappa + pcL\big[k\delta + (n-k)F(k,n)\big]$ and satisfies $k_{\mathrm{opt}} = (2n^2/\delta)^{1/3}(1 + o(1))$ for every fixed $\delta > 0$; the trapped share at the optimum is $F(k_{\mathrm{opt}}, n) = (\delta/2)^{2/3} n^{-1/3}(1 + o(1)) \to 0$ and the optimal cost is $C(k_{\mathrm{opt}}) = n\kappa + \frac{3}{2^{2/3}}\,(pcL)^{1/3} d^{2/3} n^{2/3}(1 + o(1))$.*
4. (Welfare.) *The equilibrium cost is $C(k^*) = n\kappa + nd\,(1 + O(n^{-1/2}))$ for $\delta < 1$ and $n\kappa + npcL$ for $\delta \geq 1$; the welfare loss per agent, $(C(k^*) - C(k_{\mathrm{opt}}))/n$, tends to d for $\delta < 1$ and to pcL for $\delta \geq 1$.*

Proof. (1) An A-agent facing k sinks has expected cost $\kappa + pcL\,F(k,n) \leq \kappa + pcL < pL$, so $\varnothing$ is dominated, and an O-agent has cost $\kappa + d \leq \kappa + pcLF(k-1,n) \leq \kappa + pcL < pL$ in any candidate equilibrium, so no agent chooses $\varnothing$. In a profile with k origin-checkers, an A-agent gains $pcL\,F(k,n) - d$ by switching to O and an O-agent gains $d - pcL\,F(k-1,n)$ by switching to A, because after her switch there are $k-1$ sinks and her own chain is that of a non-sink agent; neither gain is positive iff $F(k,n) \leq \delta \leq F(k-1,n)$. Strict monotonicity of F makes the solution unique for generic δ, and $k^* = 0$ iff $F(0,n) = 1 \leq \delta$. (2) By Lemma 1, $F(a\sqrt{n}, n) \to \Phi(a)$ uniformly on compact sets of a (F is monotone and Φ continuous), so $k^* = \min\{k : F(k,n) \leq \delta\}$ satisfies $k^*/\sqrt{n} \to \Phi^{-1}(\delta) = a^*$. Since $F(k^*, n) \leq \delta < F(k^* - 1, n)$ and $F(k^* - 1, n) - F(k^*, n) = O(n^{-1/2})$ (one more sink changes each term of (2) by a relative $O(1/\sqrt{n})$ at the scale $t \asymp \sqrt{n}$), $F(k^*, n) = \delta + O(n^{-1/2})$ and the cost of an A-agent tends to $\kappa + pcL\delta = \kappa + d$. The share of origin-checkers $k^*/n \to 0$, and by Proposition 2(iii) of Part I the probability of a bottom-up architecture is at most $\theta e^{1-\theta} + o(1) \to 0$ at $\theta = k^*/n$. (3) Put $k = bn^{2/3}$. By the last part of Lemma 1, $F(k,n) = n^{-1/3} b^{-2}(1 + o(1))$ uniformly for b in compact subsets of $(0, \infty)$, so $C(k) - n\kappa = pcL\,n^{2/3}\big[b\delta + b^{-2}\big](1 + o(1))$; the bracket is minimized at $b^* = (2/\delta)^{1/3}$, where it equals $b^*\delta + b^{*-2} = \frac{3}{2^{2/3}}\delta^{2/3}$. Counts outside $[\varepsilon n^{2/3}, n^{2/3}/\varepsilon]$ are suboptimal for small ε: for $k \leq \varepsilon n^{2/3}$, (2) gives $F(k,n) \geq \frac{1}{2} \cdot \min\{1, n/(4k^2)\}$ for n large (the terms with $t \leq n/(2k)$ and $t \leq \sqrt{n}$ each contribute a constant fraction), so $(n-k)F \geq c_1 n^{2/3}\varepsilon^{-2}$, while for $k \geq n^{2/3}/\varepsilon$ the cost is at least $k\delta \geq \varepsilon^{-1}\delta n^{2/3}$; both exceed the bracket's minimum times $n^{2/3}$ for ε small. Hence $k_{\mathrm{opt}}/n^{2/3} \to b^*$ and the stated cost and trapped share follow. (4) For $\delta < 1$, $C(k^*) - n\kappa = pcL[k^*\delta + (n-k^*)F(k^*,n)] = pcL[n\delta + O(\sqrt{n})] = nd + O(\sqrt{n})$ by (2); for $\delta \geq 1$, $C(0) = n\kappa + npcL$. Subtract the optimum of (3), which is $O(n^{2/3})$, and divide by n. □

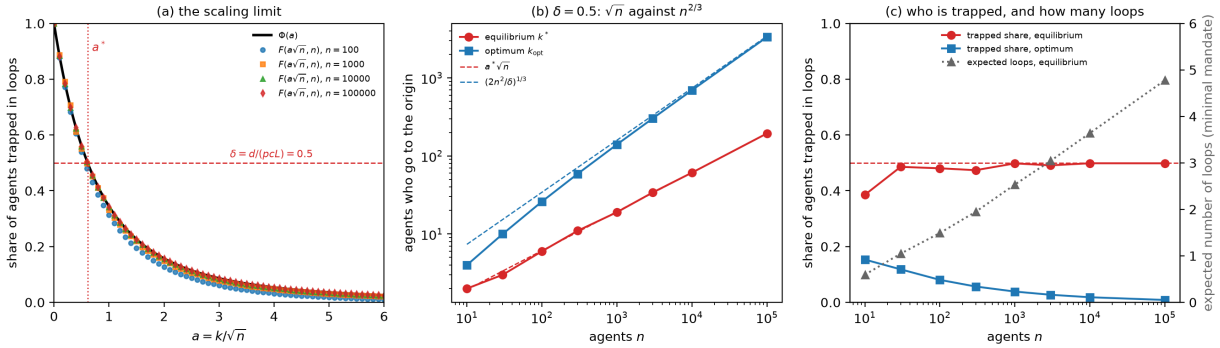

Figure 3: The network procurement game. (a) The scaling limit: exact $F(a\sqrt{n}, n)$ for $n = 10^2, \ldots, 10^5$ against $\Phi(a)$; the equilibrium of Theorem 2 sits where Φ crosses δ. (b) Equilibrium and optimal numbers of origin-checkers against n for $\delta = 0.5$ (the running example with $d = 2.5$), with the asymptotes $a^*\sqrt{n}$ and $(2n^2/\delta)^{1/3}$. (c) Share of trapped agents in equilibrium (flat at δ) and at the optimum (vanishing as $n^{-1/3}$), and the expected number of loops of the equilibrium architecture, the minimal mandate of Proposition 4, growing as $\frac{1}{2}\log n$. Generated by `network_game.py` and `make_figures.py`.

In the running example with $d = 2.5$, so $\delta = 0.5$ and $a^* = 0.612$: at $n = 100$ the equilibrium has $k^* = 6$ origin-checkers and a trapped share of 0.481, the optimum has $k_{\mathrm{opt}} = 26$ and a trapped share of 0.081, and the costs beyond $n\kappa$ are 241 against 95; at $n = 10^4$ the numbers are $k^* = 61$ against $k_{\mathrm{opt}} = 694$, trapped shares

Table 2: The running example ($p = 0.1$, $c = 0.5$, $L = 100$, $pcL = 5$, $\kappa = 3$). Costs exclude the common $n\kappa$.

	$d = 2.5$ ($\delta = 0.5$)				$d = 6$ ($\delta = 1.2$)		
n	k^*	trapped	k_{opt}	cost eq. / opt.	k^*	k_{opt}	cost eq. / opt.
10	2	0.385	4	20 / 15	0	3	50 / 27
100	6	0.481	26	241 / 95	0	18	500 / 170
1 000	19	0.498	140	2 491 / 516	0	100	5 000 / 923
10 000	61	0.499	694	24933 / 2584	0	507	50 000 / 4 624

Aicumene:2026.053v1 [econ.TH] November 2026

0.499 against 0.0182, costs 24933 against 2584, so that the equilibrium costs $nd = 25\,000$ as the theorem says and the optimum one tenth of it. With $d = 6$, the value inside the wedge of Part I ($\delta = 1.2$), nobody goes to the origin in equilibrium at any n, while the optimum sends 3 agents at $n = 10$ and 18 at $n = 100$, where it costs 170 against the equilibrium's 500 (Table 2).

Corollary 1 (The network wedge)**.** *In Γ_n^{net} the unique pure equilibrium has no origin-checker iff $d \geq pcL$, while the social optimum has at least one whenever $d < pcL\big[1 + (n-1)(1 - F(1,n))\big] = pcL\sqrt{\pi n/2}\,(1 + o(1))$. On the band*

$$pcL \;\leq\; d \;<\; pcL\sqrt{\pi n/2}\,(1 + o(1)) \tag{5}$$

every agent verifies through a loop although the group would gain from sending agents to the origin. Compared with the wedge $pcL < d < npcL$ of Part I, the upper end has shrunk from n to $\sqrt{\pi n/2}$ times the private value: the audience of one origin-checker is not the whole group but the $\sqrt{\pi n/2}$ agents whose chains happen to reach her.

Proof. The first claim is Theorem 2(1). One origin-checker beats none iff $C(1) < C(0)$, i.e. $\delta + (n-1)F(1,n) < n$, and $(n-1)(1 - F(1,n)) = \sqrt{\pi n/2}(1 + o(1))$ by Lemma 1. □

Two readings of Theorem 2. The first is the one announced in the title: the price of independence is paid by everyone and bought by almost no one. For $\delta < 1$ the marginal agent is indifferent between going to the origin and taking her chances in the network, so the expected loss of every asking agent equals the premium exactly; the group spends nd and obtains independence for a share $1 - \delta$ of its members, where $(2n^2/\delta)^{1/3}$ origin-checkers, a vanishing share, would bring the trapped share down to $n^{-1/3}$ at a cost of order $n^{2/3}$. The second reading concerns the architecture. In Part I the delegation model was a thought experiment with an exogenous share θ, and Proposition 2(iii) said that a loop-free architecture requires θ bounded away from zero. Theorem 2(2) says that no equilibrium supplies this: the share of agents who go to the origin is $a^*/\sqrt{n}$, the architecture is cyclic with probability tending to one, and the number of loops grows as $\frac{1}{2}\log n$ (Proposition 4). Loops are not an accident of a badly designed organization; they are what the procurement incentives produce.

7 Known architectures and the minimal mandate

The trapped agents of Theorem 2 are trapped because they do not know it. This section makes the architecture common knowledge and asks what changes.

Theorem 3 (Known architecture)**.** *Let the consultation partners be fixed and known before stage 1, so that the agents choose O or A on a known functional digraph D_0 (every agent has one out-arc), and let $\kappa < pL(1-c)$. Write $\mathcal{Z}$ for the set of directed cycles of D_0, which are vertex-disjoint. (i) If $d < pcL$, the pure Nash equilibria are exactly the profiles in which the set M of O-agents contains exactly one agent of every cycle of $\mathcal{Z}$ and no other agent; there are $\prod_{Z\in\mathcal{Z}} |Z|$ of them, in all of them nobody is trapped, and they coincide with the social optima. (ii) If $d > pcL$, the unique pure equilibrium is all-A, every agent is trapped, and the social optimum sends one agent to the origin on every cycle whose basin has more than δ agents.*

Proof. In a functional digraph every component contains exactly one cycle, and a set M of sinks leaves nobody trapped iff it meets every cycle, since a chain that avoids M runs into its component's cycle and around it. (i) Let M be the set of O-agents. An A-agent whose chain reaches M pays κ and would pay $\kappa + d$ as an

Aicumene:2026.053v1 [econ.TH] November 2026

O-agent, so she stays; a trapped A-agent pays $\kappa + pcL > \kappa + d$ and switches; hence in equilibrium M meets every cycle. An O-agent m pays $\kappa + d$ and would pay $\kappa + pcL \cdot \mathbf{1}[m \text{ trapped under } M \setminus \{m\}]$ as an A-agent; she stays iff she would be trapped, i.e. iff her own chain avoids $M \setminus \{m\}$, which for m on a cycle means that no other member of M lies on that cycle, and which fails for m off the cycles because her chain reaches the cycle of her component and the member of M on it. So M has exactly one agent per cycle. Conversely every such M is an equilibrium by the same inequalities. The social cost is $n\kappa + |M|d + pcL \cdot \#\{\text{trapped}\}$; leaving a cycle Z uncovered saves d and traps its basin $B_Z \supseteq Z$ with $|B_Z| \geq 2$, costing at least $2pcL > d$, so the optimum covers every cycle once. (ii) If $d > pcL$ a trapped A-agent gains at most $pcL - d < 0$ by switching, and an O-agent gains at least $d - pcL > 0$ by switching, so all-A is the unique equilibrium; the planner covers cycle Z iff $d < pcL\,|B_Z|$. □

Theorem 3 separates two causes of loops. For $d < pcL$ the loops of Theorem 2 exist only because the architecture is unknown: an agent who knew that she stands on a cycle would pay the premium, and the equilibrium with a known architecture is efficient. For $d > pcL$ the wedge of Part I returns basin by basin, $pcL < d < pcL|B_Z|$, with the basin of the loop in the role of the n receivers of one verdict. The cheapest intervention therefore differs between the two regimes: information in the first, a subsidy $s > d - pcL$ to one agent per cycle in the second, which is the subsidy of Proposition 7 of Part I applied to a set that the next proposition identifies and counts.

Proposition 4 (The minimal mandate and its size)**.** *Let D be a functional digraph on the non-sink agents with sinks absorbing. The minimum set of agents who must be made to go to the origin so that nobody is trapped is a minimum feedback vertex set of D; it consists of one agent of every directed cycle, has size $\nu(D)$, the number of cycles, and is found in linear time, although the feedback vertex set problem is NP-hard on general digraphs [30]. In the random delegation architecture with k sinks among n agents,*

$$\mathbb{E}\,\nu(D) \;=\; \sum_{t=2}^{n-k} \frac{(n-k)_t}{t\,(n-1)^t}, \qquad (m)_t = m(m-1)\cdots(m-t+1), \tag{6}$$

and $\mathbb{E}\,\nu(D) = \frac{1}{2}\log n + O(1)$ uniformly for $k \leq C\sqrt{n}$. In the equilibrium architecture of Theorem 2 the minimal mandate therefore has expected size $\frac{1}{2}\log n + O(1)$, against the $(2n^2/\delta)^{1/3}$ untargeted origin-checkers of the optimum and the $a^*\sqrt{n}$ of the equilibrium.

Proof. In a functional digraph the cycles are vertex-disjoint and every vertex leads to exactly one of them, so a set meets every cycle iff making it absorbing leaves nobody trapped, and one vertex per cycle is necessary and sufficient; the cycles are found by one pass of depth-first search. For (6), a directed t-cycle on a given ordered t-tuple of distinct non-sink agents occurs with probability $(n-1)^{-t}$, there are $(m)_t/t$ such cycles up to rotation with $m = n - k$, and self-loops are impossible; linearity of expectation gives the sum. For the asymptotics write the term as $\frac{1}{t}\prod_{j<t}(1 - \frac{k+j+1}{n-1}) \cdot (1 + O(1/n))$; for $t \leq \sqrt{n}/\log n$ and $k \leq C\sqrt{n}$ the product is $1 + o(1)$, so these terms contribute $\sum_{t=2}^{\sqrt{n}/\log n} \frac{1}{t} = \frac{1}{2}\log n - \log\log n + O(1)$; for $\sqrt{n}/\log n < t \leq \sqrt{n}\log n$ the product is $e^{-O(1)}$ at most and $\sum 1/t = 2\log\log n$, and for $t > \sqrt{n}\log n$ the product is at most $e^{-\log^2 n/2}$, negligible; the terms between contribute $O(1)$ after the $\log\log n$ corrections cancel. □

At $n = 10^4$ the equilibrium architecture of the running example has 3.6 loops on average, and 1.5 at $n = 100$ (Figure 3(c)). The contrast with Theorem 2(3) is the economic content of knowing the network: a regulator who can see D removes every loop by instructing $\frac{1}{2}\log n$ identified agents, whereas one who cannot must send $n^{2/3}$ agents to the origin and still leaves a share $n^{-1/3}$ trapped. Measurement of the architecture, who asks whom, is worth $n^{2/3}d$ in this model; in the terms of Part II, provenance is itself the public good.

8 An experiment with language-model agents: protocol

The criterion of [15] is that language-model agents reproduce an equilibrium when the decision reduces to a myopic best response over explicit payoffs, and fail to when it requires anticipating others. Γ_n^{net} separates the two in one game: the stage-3 decision is myopic, the stage-1 choice is a public-good problem with a random audience. The protocol, to be run on the cross-family stand of that paper, is as follows. (i) $n \in \{10, 20, 50\}$

Aicumene:2026.053v1 [econ.TH] November 2026

agents from at least three model families receive the same claim, the prices κ and $\kappa + d$, the parameters p, c, L, and a description of the two technologies, including the statement that a question is answered by a colleague drawn at random. (ii) Each agent chooses O or A; the architecture is drawn; verdicts are propagated by the experimenter according to the chains; each agent decides. (iii) Measured quantities: the share $\hat{\theta}$ of agents choosing O against k^*/n and k_{opt}/n; the realized trapped share against δ; the stage-3 decisions of trapped agents, which the criterion predicts to be the Bayes-myopic acceptance of Proposition 3; and, in a second treatment that reveals the architecture, the share of agents on cycles who switch to O, which Theorem 3 predicts to be one per cycle for $d < pcL$ and zero for $d > pcL$. (iv) A contamination probe: the same game with relabelled technologies and shuffled prices, to separate computation from recall of this paper. The prediction derived from the criterion is that the stage-3 cascade and the known-architecture switch are reproduced, while the $\sqrt{n}$ law of the unknown-architecture equilibrium is not, because it requires each agent to anticipate how many others will go; what the agents do instead is the quantity of interest. No results are reported here.

9 Related work

The chain process of Lemma 1 is the path of a random mapping with absorbing states; the distribution of the tail and cycle lengths of a random mapping, with the $\sqrt{\pi n/2}$ scale and the $\frac{1}{2}\log n$ cycles, is classical [31, 32], and the present contribution is to give these quantities an economic role, as the audience of an origin-checker and as the minimal mandate. Unique-sink orientations and their query complexity are the subject of [24, 25, 26, 27, 28], with the count of [11] and the acyclic count of [10, 9]; they arose as abstractions of linear complementarity and linear programming [29], and Theorem 1 reads them as the price of finding the responsible agent in an organization. Feedback vertex sets are one of the original NP-complete problems [30]; in functional digraphs the problem is trivial, which is what makes the mandate of Proposition 4 computable. Informational cascades [7, 8] and the verification game of Part II supply the dynamics of Section 5; the public-good structure is that of best-shot provision [6, 5], here consumed along arcs rather than by everyone. The closest neighbours are the public-good games on networks of [23] and the law of the few of [22], in which a few agents acquire information and the others link to them: there the links are chosen and the equilibrium is a core–periphery structure in which every link reaches a provider, whereas in Γ_n^{net} the links are drawn and may close into loops, which is why the number of providers is $\sqrt{n}$ rather than a constant and why a fixed share of agents reaches no provider at all; Theorem 3, with chosen sinks on a known map, is the point of contact between the two models; costly state verification [3] and jury aggregation with correlated votes [4] are the neighbours named in Part I, and the algorithmic agents of [21] the motivation for Section 8. The digraph of [13], three Hamiltonian cycles and no sink, is the limiting architecture of this paper, every agent trapped, and the independent check that settled it was a formal verification outside the network [14]. The structural criteria for strong intelligence of [17], the probabilistic truth of [18, 19] and the legal verification compiler of [16] are the lines of the authors' programme that this series serves.

10 Open problems

1. *Cyclic unique-sink orientations.* The share of USOs of Q_h that contain a directed cycle is 0 for $h = 2$ and 0.022 for $h = 3$; its limit as $h \to \infty$ is unknown, and the question is equivalent to a lower bound on the number of acyclic USOs sharp enough to compare with $h^{\Theta(2^h)}$. The count for $h = 4$, among 5 541 744 USOs, is within reach of a dedicated enumeration and is the next data point.
2. *Chosen partners under partial information.* Γ_n^{net} draws partners at random and Theorem 3 makes the architecture common knowledge; between them lies the game in which each agent sees her neighbourhood in a sparse consultation graph and chooses whom to ask. Its equilibria combine the sink-finding problem of Section 4 with the free-riding of Section 6.
3. *Mixed equilibria.* The symmetric mixed equilibrium of Γ_n^{net}, in which each agent goes to the origin with probability θ, has θn concentrating at k^*; its exact characterization and the comparison with learning dynamics are open, as are the no-regret limits asked for in Part I.
4. *Several origins.* With several vaults of different reliability the sink of a chain determines which provenance root is reached, and the coverage value of Part II applies to the set of roots reached by an

agent's chain; the equilibrium of that game is open.

5. *Information against welfare.* The mutual information between the truth and the verdict vector of a network, conditional on the evidence already held, orders architectures; whether that order agrees with the Bayes risk of Theorem 2 for all loss functions is not known.

11 Conclusion

The premium for independent verification is the price of reaching the origin of a record through a network of colleagues who, for the most part, have only reached one another. When the origin is reachable along a known chain, the premium is the depth, and the wedge of Part I is a band of depths and of shares. When the chain must be found under local information, the premium is a query complexity, and on the cube it outgrows any fixed stake. When the chain closes on itself, the premium is infinite, the agents on the loop confirm one another's errors without limit, and Theorem 2 says that this is the equilibrium outcome: a vanishing share of agents go to the origin, a fixed share $d/(pcL)$ verify through loops, and every agent pays the premium in expected loss without buying what it would purchase. The loops dissolve when the agents can see them, for $d < pcL$, and persist as the wedge of Part I otherwise. The cheapest remedy is therefore not always a subsidy; often it is a map of who asks whom.

Code availability and reproducibility notes

Every number in the text is produced by the scripts `network_game.py` (the exact formula (2), equilibria, optima, loop counts, and 4 000-sample Monte-Carlo checks of F and of (6) at five (n, k) pairs, all within simulation error), `uso_cycles.py` (enumeration of the orientations of Q_2 and Q_3), `depth_band.py` (the depth law and the band) and `make_figures.py`; they need only the Python standard library except for the figures, write the `numbers_*.tex` files read by this manuscript, and are available at `https://github.com/aicumene/verification-games` (directory `part3`).

References

[1] Bernadotte, A. and Menshikov, I. (2026). Multi-Agent Verification Games: Shared Sources, Crowding Out, and the Price of Independence. Part I. Aicumene Preprint 2026.047v1; SSRN working paper 7682932 (October 24, 2026). https://ssrn.com/abstract=7682932, doi:10.2139/ssrn.7682932. https://research.aicumene.com/papers/game-theory-1/.

[2] Huber, P. J. (1965). A robust version of the probability ratio test. *Annals of Mathematical Statistics* 36(6), 1753–1758. doi:10.1214/aoms/1177699803.

[3] Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory* 21(2), 265–293. doi:10.1016/0022-0531(79)90031-0.

[4] Ladha, K. K. (1992). The Condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science* 36(3), 617–634. doi:10.2307/2111584.

[5] Bergstrom, T., Blume, L. and Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics* 29(1), 25–49. doi:10.1016/0047-2727(86)90024-1.

[6] Hirshleifer, J. (1983). From weakest-link to best-shot: the voluntary provision of public goods. *Public Choice* 41(3), 371–386. doi:10.1007/BF00141070.

[7] Bikhchandani, S., Hirshleifer, D. and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100(5), 992–1026. doi:10.1086/261849.

[8] Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics* 107(3), 797–817. doi:10.2307/2118364.

[9] Stanley, R. P. (1973). Acyclic orientations of graphs. *Discrete Mathematics* 5(2), 171–178. doi:10.1016/0012-365X(73)90108-8.

Aicumene:2026.053v1 [econ.TH] November 2026

Aicumene:2026.053v1 [econ.TH] November 2026

[10] Kahale, N. and Schulman, L. J. (1996). Bounds on the chromatic polynomial and on the number of acyclic orientations of a graph. *Combinatorica* 16(3), 383–397. doi:10.1007/BF01261322.

[11] Matoušek, J. (2006). The number of unique-sink orientations of the hypercube. *Combinatorica* 26(1), 91–99. doi:10.1007/s00493-006-0007-0.

[12] Takács, L. (1990). On Cayley's formula for counting forests. *Journal of Combinatorial Theory, Series A* 53(2), 321–323. doi:10.1016/0097-3165(90)90064-4.

[13] Knuth, D. E. (2026). Claude's cycles. Note, Stanford University, 28 February 2026, revised 14 April 2026. `https://cs.stanford.edu/~knuth/papers/claude-cycles.pdf` (accessed 5 November 2026).

[14] Morrison, K. (2026). KnuthClaudeLean: a Lean formalization of the proof that Claude's construction is correct. `https://github.com/kim-em/KnuthClaudeLean/` (accessed 5 November 2026).

[15] Bernadotte, A. and Menshikov, I. (2026). When do language models reach Nash? Heterogeneous cross-family LLM agents in classic economic games. SSRN working paper 7637178 (June 2026). `https://ssrn.com/abstract=7637178`, `https://research.aicumene.com/papers/llm-nash/`, doi:10.2139/ssrn.7637178.

[16] Bernadotte, A. and Menshikov, I. (2026). Verifiable AI for Law: A Legal Verifiability Compiler with a Rule/Judgment Firewall. Aicumene Preprint 2026.040. doi:10.13140/RG.2.2.18174.75841; ResearchGate publication 416039549, `https://www.researchgate.net/publication/416039549`. `https://research.aicumene.com/papers/verifiable-law/`

[17] Bernadotte, A. and Gromov, V. A. (2025). What Is Strong AI? Intrinsic Structural Criteria for Strong Intelligence Based on Dimensionality and Topology. Aicumene Preprint 2025.043. doi:10.13140/RG.2.2.12211.59681. `https://research.aicumene.com/papers/StrongAI/`

[18] Bernadotte, A. (2026). The April Theses of the New Cognitive Paradigm: From Imitation to Structural Measurability. Aicumene Preprint 2026.046 (v1, April 2026; v2, October 2026). doi:10.13140/RG.2.2.20527.50087, `https://research.aicumene.com/papers/NCP/`.

[19] Bernadotte, A. and Menshikov, I. (2026). Safe AI by Design: Probabilistic Truth, Dynamic Structural Safety, and a Quantitative Harm Model for Agentic Generative AI. Aicumene Preprint 2026.050v1, `https://research.aicumene.com/papers/Safe-by-Design/`, SSRN forthcoming.

[20] Bernadotte, A. and Menshikov, I. (2026). Multi-Agent Verification Games II: Information Entrenchment and Competitive Underprovision of Independent Evidence. Part II. Aicumene Preprint 2026.048v1; SSRN working paper 7747018 (October 24, 2026). `https://ssrn.com/abstract=7747018`, doi:10.2139/ssrn.7747018. `https://research.aicumene.com/papers/game-theory-2/`.

[21] Calvano, E., Calzolari, G., Denicolò, V. and Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110(10), 3267–3297. doi:10.1257/aer.20190623.

[22] Galeotti, A. and Goyal, S. (2010). The law of the few. *American Economic Review* 100(4), 1468–1492. doi:10.1257/aer.100.4.1468.

[23] Bramoullé, Y. and Kranton, R. (2007). Public goods in networks. *Journal of Economic Theory* 135(1), 478–494. doi:10.1016/j.jet.2006.06.006.

[24] Szabó, T. and Welzl, E. (2001). Unique sink orientations of cubes. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, 547–555. doi:10.1109/SFCS.2001.959931.

[25] Schurr, I. and Szabó, T. (2004). Finding the sink takes some time: an almost quadratic lower bound for finding the sink of unique sink oriented cubes. *Discrete & Computational Geometry* 31(4), 627–642. doi:10.1007/s00454-003-0813-8.

[26] Gärtner, B. (2002). The Random-Facet simplex algorithm on combinatorial cubes. *Random Structures & Algorithms* 20(3), 353–381. doi:10.1002/rsa.10034.

[27] Matoušek, J. and Szabó, T. (2006). RANDOM EDGE can be exponential on abstract cubes. *Advances in Mathematics* 204(1), 262–277. doi:10.1016/j.aim.2005.05.021.

[28] Gärtner, B. and Schurr, I. (2006). Linear programming and unique sink orientations. In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms*, 749–757. doi:10.1145/1109557.1109639.

[29] Stickney, A. and Watson, L. (1978). Digraph models of Bard-type algorithms for the linear complementarity problem. *Mathematics of Operations Research* 3(4), 322–333. doi:10.1287/moor.3.4.322.

[30] Karp, R. M. (1972). Reducibility among combinatorial problems. In *Complexity of Computer Computations*, 85–103. Plenum. doi:10.1007/978-1-4684-2001-2_9.

[31] Harris, B. (1960). Probability distributions related to random mappings. *Annals of Mathematical Statistics* 31(4), 1045–1062. doi:10.1214/aoms/1177705677.

[32] Flajolet, P. and Odlyzko, A. M. (1990). Random mapping statistics. In *Advances in Cryptology — EUROCRYPT '89*, Lecture Notes in Computer Science 434, 329–354. Springer. doi:10.1007/3-540-46885-4_34.

Aicumene:2026.053v1 [econ.TH] November 2026