Aicumene:2026.048v1 [econ.TH] October 2026

Multi-Agent Verification Games II: Information Entrenchment and Competitive Underprovision of Independent Evidence

Aicumene Preprint

Alexandra Bernadotte
Aicumene*

Ivan Menshikov
Aicumene

October 2026 — working paper, Part II (SSRN preprint)

Abstract

A publicly shared verification result may induce wide reliance while preserving an error that originates in a source common to the proposer and the verifier. We study the demand and the supply of source independence. On the demand side, in a sequential game in which receivers observe earlier decisions and hold private independent signals, a shared "ok" verdict starts an immediate and permanent cascade whenever a private signal's likelihood ratio is below an explicit threshold $\Lambda = (1-p)\ell/(pcL)$: every receiver accepts a common-source error with probability pc for every horizon, and n private signals whose pooled content would identify the state go unused; above the threshold the outcome is decided by the first receiver's signal alone. In a band of prices for an independent check nobody buys one although the first receiver's purchase would be efficient. On the supply side the verdict is a non-excludable public good. With observable, contractible quality and posted-price competition under an explicit one-certificate procurement protocol, every pure subgame-perfect equilibrium implements the copied-source technology at cost whenever the cost premium of independence exceeds one receiver's private value, although independence is efficient whenever the premium is below n times that value; a monopolist faces the same comparison. A hidden post-contract source choice removes independence even below the private threshold, and an audit-and-sanction threshold restores it. Removing the one-certificate rule, with any number of conditionally independent provenance roots the residual false-accept probability is the product of the roots' failure shares, and the private–social wedge recurs at every diversity margin with value $pL(1-c)c^k$, shrinking geometrically. When sources copy one another, the all-ok probability is the mass of the common ancestors of the purchased sources: the value of a set of certificates is a submodular coverage function, a copy of copies adds nothing, disjoint provenance catches everything that is left, no number of certificates from one root lowers the floor, and independent contamination is neither the best nor the worst case. The four results share one structure: the private value of source independence is v, its social value is nv, and independence is under-provided exactly on the band (v, nv), whose width grows with the number of decisions that rely on one public verdict and vanishes with the common-source share. We then remove the one-certificate restriction: distinct independent roots generate a multiplicative residual error probability, and the privately supported number of origins is weakly below the efficient number. Price competition is addressed explicitly through a cost-price benchmark; the unrestricted simultaneous price–quality supply equilibrium is not claimed solved. The results are conditional, and finite examples are checked by exhaustive enumeration.

Keywords: costly verification; public information; correlated signals; sequential learning; hidden action; moral hazard; source provenance; information economics.
JEL: D82, D83, D86, H41.

*Aicumene: `https://aicumene.com/`. Corresponding author: bernadotte@aicumene.com.
© 2026 Alexandra Bernadotte and Ivan Menshikov. Licensed under CC BY-NC-ND 4.0.

Aicumene:2026.048v1 [econ.TH] October 2026

1 Introduction: why duplicated checking does not create independent evidence

Checking a claim repeatedly and checking it from independent origins are distinct economic activities. A registry copied by several offices, a legal authority reproduced in several databases, and two language-model agents drawing on the same stale source can generate multiple confirmations of one error. The point is not that public information is inherently harmful. Its provenance matters because public circulation magnifies the number of decisions that rely on a single informational root, while no additional independent evidence is necessarily produced.

The conceptual difficulty is thus *not* solving a larger payoff matrix. It is choosing an information structure in which the replication of signals, the sequential diffusion of actions, and the supply of evidence independence can be separated. We ask: can downstream action histories correct a common-source mistake? And will competition in verification services create independent provenance if that independence is an unobservable, costly choice of the provider?

Morris and Shin's study of the social value of public information emphasizes that publicly observed signals affect coordination and welfare under strategic complementarities [4]. Our agents need not have a coordination motive in payoffs: the distinct mechanism here is *a failure of statistical identification caused by a shared origin*. In particular, public dissemination of a verdict is not statistically equivalent to gathering another independent signal. This distinction is not an argument against public transparency; the interpretation of the Morris–Shin result itself was debated by Svensson and by Morris, Shin, and Tong [5, 6].

We present the one-certificate benchmark alongside a new multi-certificate game that distinguishes independent origins from repeated copies. We retain four earlier results and one shared structure. On the demand side, an indistinguishability benchmark extends the information floor from signals to arbitrary public action histories, and the entrenchment theorem adds private independent signals: a shared verdict starts an immediate cascade below an explicit precision threshold, so public reliance never uses the private information that could correct the error, and in a band of prices nobody buys the independent check that would. On the supply side, the verdict is a non-excludable public good: under an explicit procurement protocol with observable quality, every pure subgame-perfect equilibrium of posted-price competition supplies the copied source whenever independence costs more than one receiver's private value, although it is efficient up to n times that value; a hidden post-contract source choice and an audit threshold form the second layer. The four results share one arithmetic, the independence wedge (v, nv). A theorem on correlated roots replaces the number of certificates by the geometry of their ancestry. None of them is offered as the first discovery of informational cascades or of hidden-quality market failure.

2 Primitive information structure and benchmark

A proposition is either correct (R) or wrong (W), with $\Pr(W) = p \in (0, 1)$. Conditional on W, a fraction $c \in [0, 1]$ arises from a common faulty origin E; the rest is idiosyncratic. A perfect shared-source verification produces "ok" in states R and (W, E), and "flag" in (W, E^c). A perfect independent-source verification perfectly distinguishes R from W. Public acceptance of W entails loss $L > 0$ per receiver; rejection of R entails loss $\ell > 0$. We assume

$$pL < (1 - p)\ell, \tag{1}$$

so accepting an unchecked claim is Bayes-optimal. With perfect signals, the expected receiver loss after a shared check is pcL and after an independent check is zero. The benchmark in [1] derives an information floor and a one-period crowding-out region for alternative purchases. We use its primitives, but change the timing and introduce supplier effort instead of relabeling consumer technology choice as endogenous quality.

Lemma 1 (Common-source likelihood equivalence)**.** *For any collection of conditionally independent shared-source checks, and any additional public randomization that does not depend directly on the state, the joint observations have equal laws under R and (W, E).*

Proof. Each shared check has the same conditional signal law in both events by definition. Products and common randomization preserve equality of the induced joint laws. □

Aicumene:2026.048v1 [econ.TH] October 2026

3 Sequential reliance and the entrenchment theorem

There are $n \geq 2$ receivers arriving in order. Receiver i sees the public history H_i of previous purchases, available verdicts, and public accept/reject actions. The receiver may procure another shared-source check, procure nothing, and finally select $a_i \in \{0, 1\}$, where 1 means acceptance. Behavioral strategies $\sigma_i(H_i)$ can randomize and are allowed to depend on the entire public history. No receiver observes a private signal sourced independently of the original Σ. All endogenous randomizations have the same law across states. The restriction deliberately isolates *sequential copying without independent information*; in this section procurement incentives are not needed for the information bound.

Let $A_n = \{a_1 = \cdots = a_n = 1\}$ be collective acceptance. Write $\varepsilon_n = \Pr(A_n^c \mid R)$ for the probability that at least one receiver refuses a correct claim.

Theorem 1 (Sequential indistinguishability: the benchmark)**.** *For every finite horizon n and every behavioral strategy profile using only the allowed shared-source signals and public actions,*

$$\Pr(W, E, A_n) = pc \Pr(A_n \mid R) = pc(1 - \varepsilon_n). \tag{2}$$

Consequently, if a family of protocols has $\varepsilon_n \to 0$, then $\liminf_{n\to\infty} \Pr(W, A_n) \geq pc$. Adding any number of downstream public actions or source-sharing checks cannot lower this bound without reducing correct-claim collective acceptance.

Proof. Induct on the receiver index. The initial histories have identical conditional laws under R and (W, E). Suppose the public history prior to receiver i has the same distribution in these two events. Conditional on that history, receiver i applies the same randomized strategy, and any new shared-source check has the same conditional distribution under both events by Lemma 1. Thus the next public history also has the same law. The induction gives $\Pr(A_n \mid W, E) = \Pr(A_n \mid R)$. Multiplying by $\Pr(W, E) = pc$ proves the identity. The lower bound follows because $\Pr(W, A_n) \geq \Pr(W, E, A_n)$. □

The statement deliberately avoids a universal lower bound of pc on unconditional unanimous acceptance: a protocol that rejects all propositions has $\varepsilon_n = 1$ and makes the lower bound zero. The theorem instead identifies an *accuracy–entrenchment frontier*, indexed by the acceptance probability on correct claims. It holds for every strategy profile and therefore has no strategic content; its role is to isolate what the next result adds, namely that a fresh source available to every receiver need not break the premise.

Private independent signals

We now give each receiver what the benchmark withholds. The shared check has been bought once and its verdict is public. Receiver i additionally observes a private signal $s_i \in \{\mathsf{ok}, \mathsf{flag}\}$ from an independent source, $\Pr(s_i = \mathsf{flag} \mid W) = \Pr(s_i = \mathsf{ok} \mid R) = \pi \in (\frac{1}{2}, 1)$, independent of E and across receivers conditional on the state. Write $\lambda = \pi/(1 - \pi) > 1$ for its likelihood ratio and

$$\Lambda = \frac{(1 - p)\ell}{pcL} > 1 \tag{3}$$

for the likelihood ratio a single signal needs in order to overturn acceptance after a shared ok. Receiver i sees the public verdict, the actions $a_1, \ldots, a_{i-1}$ and s_i, and chooses a_i. There are no payoff externalities, so sequential rationality is Bayesian myopia given correct beliefs about predecessors' strategies; we exclude the tie $\lambda = \Lambda$.

Theorem 2 (Entrenchment with private signals)**.** *Let the shared verdict be ok; on (W, E^c) it is flag and every receiver rejects.*

1. *If $\lambda < \Lambda$, then in every PBE, at all histories reached with positive probability after the shared ok, every receiver accepts regardless of her signal: $\Pr(A_n \mid R) = 1$ and $\Pr(W, A_n) = pc$ for every n. The n private signals, whose pooled content would identify the state with probability tending to one, are never used.*
2. *If $\lambda > \Lambda$, then in every PBE receiver 1 follows her signal and, conditional on her acceptance, every later receiver accepts regardless of her signal along the reached history; hence $A_n = \{s_1 = \mathsf{ok}\}$, $\Pr(A_n \mid R) = \pi$, $\Pr(A_n \mid W, E) = 1 - \pi$ and $\Pr(W, A_n) = pc(1 - \pi)$ for every n.*

Aicumene:2026.048v1 [econ.TH] October 2026

In both cases $\Pr(W, A_n) \geq pc(1-\pi)$ *for every horizon: no length of public reliance corrects a common-source error better than the first receiver's single private signal. Without the shared verdict a private signal is used iff* $\lambda > \Lambda_0 = (1-p)\ell/(pL) = c\Lambda$*; the public shared verdict raises the bar for private information by the factor* $1/c$*.*

Proof. Write O_i for the public odds of W against R before receiver i moves. After the shared ok, $O_1 = pc/(1-p)$, and $O_1 < \ell/L$ by (1). Receiver i rejects iff her posterior odds exceed ℓ/L, i.e. iff $O_i\lambda > \ell/L$ after flag and iff $O_i/\lambda > \ell/L$ after ok. (1) If $\lambda < \Lambda$ then $O_1\lambda < \ell/L$, so receiver 1 accepts after both signals; her action is then the same under R and (W, E), so $O_2 = O_1$ and the argument repeats. (2) If $\lambda > \Lambda$ then $O_1\lambda > \ell/L > O_1/\lambda$, so receiver 1 follows her signal and her action reveals it. After $a_1 = 1$ the odds are $O_2 = O_1/\lambda$, and $O_2\lambda = O_1 < \ell/L$: receiver 2 accepts after both signals, her action is uninformative, and by induction so does everyone after her. After $a_1 = 0$ the event A_n has already failed. Hence $A_n = \{s_1 = \mathsf{ok}\}$, which has probability π under R and $1-\pi$ under (W, E). The claim about Λ_0 repeats the computation with prior odds $p/(1-p)$ in place of O_1. □

At $p = 0.1$, $c = 0.5$ and $L = \ell$ the threshold is $\Lambda = 18$: after one shared ok, a private independent signal that is right with probability below $\pi^* = \Lambda/(1+\Lambda) = 0.947$ is ignored by everyone. This is the informational cascade of [10, 11] with one difference: there a cascade forms after a few informative actions, here the public shared verdict starts it at the first receiver, so the public history never carries information about the one event the private signals could resolve. Figure 1 shows $\Pr(W, A_n)$ against π: flat at pc below π^*, equal to $pc(1-\pi)$ above it.

Corollary 1 (Sequential crowding out)**.** *Let* $\lambda < \Lambda$ *and let each receiver, at her turn, be able to buy a perfect independent check at price* κ_I *whose verdict becomes public. Write* $v = L\Pr(W \mid \mathsf{ok}, \mathsf{flag}) = pc\lambda L/(pc\lambda + 1 - p)$ *for the value of such a check to a receiver who has seen a shared* ok *and a private* flag*. On the band*

$$v < \kappa_I < n\,v \tag{4}$$

nobody buys in any perfect Bayesian equilibrium, although a purchase by receiver 1 *upon a* flag *signal would be efficient; the common-source error is then accepted by all* n *receivers with probability* pc *while* n *private signals and an affordable independent check go unused. For* $\kappa_I < v$ *receiver* 1 *buys after a* flag *and the error is caught at the first step with probability* π*.*

Proof. A receiver who has not bought accepts (Theorem 2(1)); her expected loss is L times her posterior probability of W, which is $pc\lambda/(pc\lambda + 1 - p)$ after flag and smaller after ok; a check reduces it to zero and costs κ_I, so no receiver buys iff $\kappa_I > v$. Since predecessors' actions are uninformative, every receiver faces the same decision, and later purchases do not help a receiver who has already acted. A purchase by receiver 1 after flag makes the truth public for all n receivers, with expected gain nv. □

4 A verification market with observable quality

We now make the verifier's *own source choice* endogenous. Providers sell a public verdict used by $n \geq 2$ identical receivers. Technology C costs a provider $K_C > 0$ and inherits common-source mistakes with probability c_C; technology I costs $K_I = K_C + \Delta K$, $\Delta K > 0$, and inherits them with probability c_I, $0 \leq c_I < c_C < 1$; full independence is $c_I = 0$. Certificates carry no own error, so a receiver's expected decision loss is pc_tL after a certificate of technology t and pL with none. In this section source quality is observable, verifiable and contractible; Section 6 removes that. The verdict, once registered, is public and non-excludable: a receiver who has not paid uses it all the same.

Protocol. There are $m \geq 2$ symmetric firms. Every firm simultaneously posts one offer (t_j, P_j), $t_j \in \{C, I\}$, $P_j \geq 0$. Receivers observe the menu and simultaneously choose no order or an order from one firm. A public registrar admits at most one certificate per matter and implements the order of the lowest-index ordering receiver, breaking ties among firms by index; only the implemented order is charged, at its posted price, and the implemented firm earns $P_j - K_{t_j}$. The one-certificate limit is a market rule, not a consequence of information technology; permitting several certificates creates a different game. Write $g_j = P_j + pc_{t_j}L$ for the generalized cost of offer j to the receiver who pays for it, and $g_* = \min_j g_j$.

Aicumene:2026.048v1 [econ.TH] October 2026

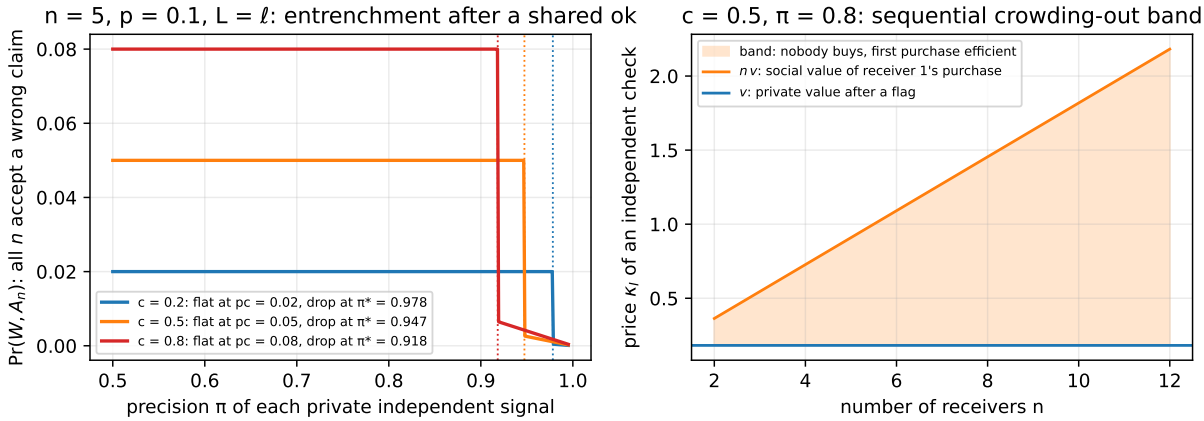

Figure 1: Left: probability that all $n = 5$ receivers accept a wrong claim after a shared ok, as a function of the precision of each private independent signal, from exact enumeration of the equilibrium path ($p = 0.1$, $L = \ell$). Dotted lines: $\pi^* = \Lambda/(1+\Lambda)$. Right: the band (4) of prices of an independent check in which nobody buys although the first receiver's purchase is efficient. Generated by `seq_model.py`.

Continuation selection. After every menu: if $g_* < pL$, receiver 1 orders from a firm attaining g_* (ties by index) and receivers $2, \dots, n$ abstain; if $g_* \geq pL$, nobody orders. This is a Nash equilibrium of the funding subgame: receiver 1 takes her privately optimal action; given that she orders, another receiver's order is not implemented and changes nothing; given that nobody orders, an order at $g_* \geq pL$ does not pay. Other funding equilibria exist, since any receiver can be the one who pays; we fix receiver 1 as the payer and analyze pure-strategy subgame-perfect equilibria of the two-stage game under this selection.

Theorem 3 (Competitive sourcing failure with observable quality)**.** *Suppose* $pL < (1-p)\ell$ *and*

$$0 < K_C < pL(1 - c_C), \qquad pL(c_C - c_I) < \Delta K < npL(c_C - c_I). \tag{5}$$

Then a pure subgame-perfect equilibrium exists, and in every pure subgame-perfect equilibrium the implemented certificate is of technology C *at price* K_C*, bought by receiver* 1*, and every firm earns zero profit. The cost-minimizing allocation of at most one certificate instead buys technology* I*, and the equilibrium social excess cost is exactly*

$$\mathcal{W} = npL(c_C - c_I) - \Delta K > 0. \tag{6}$$

The conclusion holds for every $m \geq 2$*: more price-competing firms, zero markups and verifiable source quality do not implement efficient independence.*

Proof. Existence. Let every firm post (C, K_C). Then $g_* = K_C + pc_C L < pL$ by the first condition, receiver 1 buys from firm 1, and all firms earn zero. A firm raising its price loses the order; lowering it sells at a loss; switching to I at a price P wins only if $P + pc_I L \leq K_C + pc_C L$, i.e. $P \leq K_C + pL(c_C - c_I) < K_I$ by the first inequality of the premium band, which is a loss. No deviation is profitable.

Characterization. Take any pure profile and its outcome under the selection. If no certificate is implemented, every $g_j \geq pL$, and any firm can post (C, P') with $K_C < P' < pL - pc_C L$, a nonempty interval by the first condition, win and earn $P' - K_C > 0$. If a certificate of technology I is implemented at price P, the winner earns $P - K_I$, so $P \geq K_I$, since otherwise withdrawing to a losing price is profitable; then $g_* = P + pc_I L \geq K_I + pc_I L > K_C + pc_C L$ by the first inequality of the band, and any other firm can post (C, P') with $K_C < P' < g_* - pc_C L$, win strictly and earn $P' - K_C > 0$. If a certificate of technology C is implemented at a price $P \neq K_C$: for $P < K_C$ the winner withdraws, and for $P > K_C$ any other firm posts (C, P') with $K_C < P' < P$, wins and earns $P' - K_C > 0$. Hence every pure subgame-perfect equilibrium implements C at K_C.

Welfare. Transfers cancel. One C certificate costs $K_C + npc_C L$, one I certificate $K_I + npc_I L$, and none npL. The second inequality of the band makes I cheaper than C, and $K_I + npc_I L < K_C + npc_C L < npL$ because $K_C < pL(1 - c_C) \leq npL(1 - c_C)$. Subtraction gives (6). □

Remark 1 (Monopoly). With $m = 1$ and the same protocol, the tie $g_* = pL$ resolved in favour of ordering, the monopolist's best price for technology t is $pL(1 - c_t)$, one receiver's private value; its profit is $pL(1 - c_t) - K_t$ and it chooses I iff $\Delta K < pL(c_C - c_I)$, the same comparison as under competition, with the surplus moved from the receiver to the firm. The failure is non-excludability, not market structure.

Corollary 2 (The private and social independence premiums)**.** *Independence is supplied iff* $\Delta K < pL(c_C - c_I)$ *and efficient iff* $\Delta K < npL(c_C - c_I)$*. The band on which it is efficient but not supplied has width* $(n - 1)pL(c_C - c_I)$*, the external benefit of independence to the* $n - 1$ *receivers who do not pay, and it is nonempty iff* $c_C > c_I$*. With* $c_I = 0$ *and* $\Delta K = d$ *it is the procurement wedge of Appendix A, now derived from the supply side.*

Example. Let $p = .1$, $L = \ell = 1$, $n = 4$, $c_C = .6$, $c_I = .1$, $K_C = .02$, $\Delta K = .09$, and $m = 3$. The private gain from independent provenance is .05 against its extra resource cost .09, while the collective gain is .20. Firms competitively post a correlated check priced .02; the social cost is $.02 + 4(.06) = .26$, against $.11 + 4(.01) = .15$ with independent provenance and .4 with no verification. The equilibrium welfare loss is .11 per matter.

Scope and limits. The theorem characterizes pure subgame-perfect equilibria under the stated protocol and continuation selection, not all equilibria of all procurement games. The registrar's one-certificate limit, the payer selection, price-taking Bayes-optimal receivers, firm cost symmetry and an observable enforceable quality contract are explicit assumptions. Collective purchasing, advance cost sharing, exclusive access, other registrar rules, differentiated buyers or several complementary evidence origins may reverse the conclusion. The theorem isolates underprovision caused by non-excludable public information after the hidden-action friction has been removed; the appendix's subsidy and a buyers' consortium are its remedies.

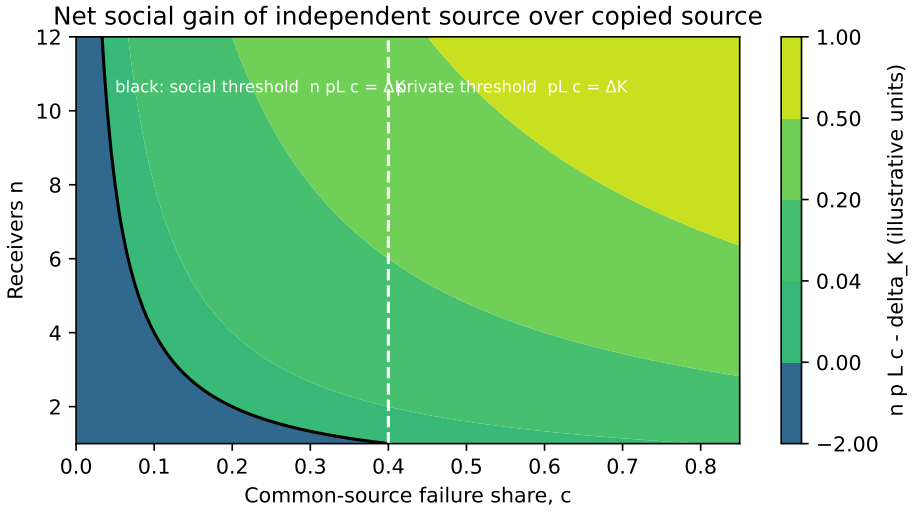

Figure 2: Net social gain $npLc - \Delta K$ of independently sourced verification over copied-source verification ($c_I = 0$) as the number of receivers and the shared-source error share vary; $p = .1$, $L = 1$, $\Delta K = .04$. Black line: social threshold $npLc = \Delta K$; white dashed line: private threshold $pLc = \Delta K$ of Theorem 3. Between them independence is efficient but not supplied. Illustrative parameters, not field measurements.

5 Unrestricted procurement of evidence origins

We now *remove* the registrar's one-certificate restriction. The following extension uses a distinct and explicit joint law of source failures; it cannot be inferred from marginal error shares c_C, c_I alone. This separation is essential: two certificates can be institutionally distinct and yet statistically identical.

Aicumene:2026.048v1 [econ.TH] October 2026

Aicumene:2026.048v1 [econ.TH] October 2026

Sources and timing. There is a finite set J of potential certificates. Certificate j has an observable, verifiable provenance root $r(j)$, price $P_j > 0$, and a source-failure indicator $F_{r(j)}$ conditional on W. A purchased certificate returns flag iff its root does not fail. Under R it always returns ok. All certificates from a given root have the same signal; distinct roots fail independently conditional on W, with $\Pr(F_r = 1 \mid W) = c_r \in [0, 1]$. A receiver can purchase any subset of as-yet unpurchased certificates, pays for its own purchases, and all verdicts are publicly observed. At the procurement stage n receivers simultaneously submit finite sets of orders; every certificate is supplied and charged at most once, with buyer-index priority for duplicate orders. Receiver decisions follow Bayes' rule after all reports. Purchasing the same root a second time has no value and costs strictly more, so each equilibrium and every efficient allocation use at most one certificate per root. Thus there is *no cap on total certificates*; only redundant copies are eliminated endogenously.

For a set T of distinct roots define $C(T) = \prod_{r \in T} c_r$, with $C(\varnothing) = 1$. Under the maintained accept-by-default condition $pL < (1-p)\ell$, accept if and only if every purchased certificate reports ok. Indeed the posterior odds of W after all ok are $pC(T)/(1-p)$; after any flag the claim is certainly wrong. Consequently each receiver's expected decision loss is $pLC(T)$.

Theorem 4 (Multiplicative information floor and marginal value)**.** *For any collection T of distinct provenance roots, the residual false-accept probability under the Bayes rule is*

$$F_A(T) = p \prod_{r \in T} c_r, \qquad F_R(T) = 0. \tag{7}$$

The private gain of buying an additional independent root $j \notin T$ is

$$b_j(T) = pL(1 - c_j) \prod_{r \in T} c_r, \tag{8}$$

and its aggregate benefit for n receivers is $nb_j(T)$. Consequently an additional certificate is individually worthwhile at posted price P_j precisely when $P_j < b_j(T)$, whereas it is socially worthwhile at real cost K_j precisely when $K_j < nb_j(T)$, provided the comparison holds other certificates fixed.

Proof. Conditional on W, all distinct roots must fail for every certificate to report ok. Conditional independence gives $C(T)$. On R all reports are ok. Bayes acceptance follows from $pLC(T) \le pL < (1-p)\ell$. The difference of individual decision losses between T and $T \cup \{j\}$ is $pLC(T)(1 - c_j)$; summing over n receivers yields the social comparison. □

Entry versus replacement. Take one copied-source root C with failure probability c_C and price P_C, and one independent root I with $c_I < c_C$ and price P_I; $n \ge 2$ receivers may order any subset of the two, duplicate orders are assigned by buyer index and charged once, and verdicts are public. Two margins now coexist: a receiver who pays nothing would *add* I to an installed C if $P_I < pLc_C(1 - c_I)$, while the payer of C would *replace* it if $P_I - P_C < pL(c_C - c_I)$.

Proposition 1 (Two heterogeneous roots)**.** *Away from equalities, the pure-equilibrium outcomes, described by the set of purchased roots, are exactly:*

1. *$\varnothing$ iff $P_C > pL(1 - c_C)$ and $P_I > pL(1 - c_I)$;*
2. *$\{C\}$ iff $P_C < pL(1 - c_C)$, $P_I > pLc_C(1 - c_I)$ and $P_I - P_C > pL(c_C - c_I)$;*
3. *$\{I\}$ iff $P_I < pL(1 - c_I)$, $P_C > pLc_I(1 - c_C)$ and $P_I - P_C < pL(c_C - c_I)$;*
4. *$\{C, I\}$ iff $P_C < pLc_I(1 - c_C)$ and $P_I < pLc_C(1 - c_I)$.*

In particular $\{C\}$ is an equilibrium outcome while adding I is efficient exactly when, in addition to the conditions in (2), $pLc_C(1 - c_I) < P_I < npLc_C(1 - c_I)$ and P_I is the real cost of I: the band (v, nv) with $v = pLc_C(1 - c_I)$. The one-certificate protocol of Section 4 forbids the outcome $\{C, I\}$ and thereby replaces this addition margin by the replacement margin $pL(c_C - c_I)$.

Proof. A receiver's cost is $pL\,C(T)$ plus her own payments, where T is the set of purchased roots. Ordering a root already bought by a higher-index receiver transfers its payment to the deviator without changing T, so it is never profitable; every other unilateral deviation changes T by adding or dropping the deviator's own roots. For $\varnothing$ the only deviations add C, I or both, with gains $pL(1-c_C) - P_C$, $pL(1-c_I) - P_I$ and $pL(1-c_Cc_I) - P_C - P_I$;

the third is negative when the first two are, since $(1-c_C)+(1-c_I) \geq 1-c_Cc_I$. For $\{C\}$ the payer may drop C, add I or switch to I, and a non-payer may add I; the three listed inequalities are exactly these four conditions. $\{I\}$ is symmetric. For $\{C, I\}$ each root's owner may drop it, with the two listed conditions; if one receiver owns both, dropping both is unprofitable because $pL(1-c_Cc_I) \geq pLc_I(1-c_C)+pLc_C(1-c_I)$. The efficiency statement compares n times the loss reduction $pLc_C(1-c_I)$ with the real cost of I. □

Theorem 5 (Exact symmetric multi-origin procurement equilibrium)**.** *Suppose M independent, exchangeable origins are offered at the fixed, common posted price $K > 0$, each with $0 < c < 1$; buyers can purchase any available certificate, every purchase creates a public signal, and duplicate orders are assigned at most once. For $0 \leq k \leq M$, a pure procurement Nash equilibrium with exactly k purchased roots exists if and only if*

$$K \geq pL(1-c), \qquad k = 0, \tag{9}$$

$$pL(1-c)c^k \leq K \leq pL(1-c)c^{k-1}, \qquad 0 < k < M, \tag{10}$$

$$K \leq pL(1-c)c^{M-1}, \qquad k = M. \tag{11}$$

The socially optimal number $k^* \in \{0, \ldots, M\}$ minimizes $kK + npLc^k$ and satisfies the same threshold conditions with pL replaced by npL. Away from boundary equalities the equilibrium number k^{NE} and the optimum are unique and $k^{NE} \leq k^*$.

Proof. At a configuration with k distinct roots, an owner who cancels one certificate raises her own expected loss by $pL(1-c)c^{k-1}$ and saves K, and a receiver who buys a $(k+1)$st root lowers her loss by $pL(1-c)c^k$ and pays K; the stated inequalities are therefore necessary. They are also sufficient against multi-unit deviations. By exchangeability a receiver's cost depends only on the total number of purchased roots and on the number she pays for, so a deviation that drops d of her certificates and adds a new ones acts as a net change of the count by $a-d$. Adding a roots to k changes her cost by $aK - pL(1-c)c^k \sum_{i<a} c^i \geq pL(1-c)c^k\big(a - \sum_{i<a} c^i\big) \geq 0$ under the lower bound; dropping d changes it by $pL(1-c)c^{k-d}\sum_{i<d} c^i - dK \geq pL(1-c)c^{k-d}\big(\sum_{i<d} c^i - dc^{d-1}\big) \geq 0$ under the upper bound, since $c^i \geq c^{d-1}$ for $i \leq d-1$. Ordering a root owned by a higher-index receiver only transfers its payment. The social cost $kK + npLc^k$ has increments $K - npL(1-c)c^k$, increasing in k because $c < 1$, which gives the optimum by the same thresholds with npL in place of pL; since $npL(1-c)c^k \geq pL(1-c)c^k$, $k^{NE} \leq k^*$. Away from equalities the intervals in K are disjoint, so both counts are unique. □

Corollary 3 (Independent sources enter and scale)**.** *At fixed positive price K the first independent root is purchased in a pure equilibrium if $K < pL(1-c)$; the socially desirable first root exists if $K < npL(1-c)$. For interior k, a further $(k+1)$st root is privately profitable only if $K < pL(1-c)c^k$, while socially profitable if $K < npL(1-c)c^k$. Thus the private–social entry wedge recurs at every diversity margin, shrinking geometrically as the number of independent roots increases.*

Supply-side interpretation and a precise limit. If each provenance root can be supplied by at least two technologically identical, price-competing vendors with constant per-certificate cost K, and firms compete *only within their assigned provenance class*, Bertrand competition implements price K in each class. Theorem 5 then describes the resulting buyers' procurement subgame and, together with these restricted vendor games, a subgame-perfect outcome. If vendors can strategically create or switch roots, price discriminate, bundle checks, or condition service on a buyer's order, a separate multi-product price–quality equilibrium analysis is required. In particular, the v13 claim that every supplier posts the copied technology cannot simply be transferred to the unrestricted certificate market. With any number of conditionally independent origins, the correct replacement for the old additive residual-error model is equation (7).

Numerical illustration. Take $p = 0.1$, $L = 1$, $n = 4$, $c = 0.5$, $K = 0.006$ and $M = 8$. Then $k^{NE} = 4$ but $k^* = 6$: the expected total resource loss is 0.049 at equilibrium against 0.04225 at the optimum, a gap of 0.00675 per matter. These are illustrative calculations, not empirical estimates; the numbers are produced by `multi_check.py`.

Aicumene:2026.048v1 [econ.TH] October 2026

Aicumene:2026.048v1 [econ.TH] October 2026

Correlated roots: sources that copy one another

The product formula describes roots contaminated independently. Copies are not like that: a registry that copies another inherits exactly the errors of its parent, and one error has one origin. Let the sources be nodes of a directed acyclic graph, as in the source-tree proposition of Part I [1]: an edge $\sigma \to \sigma'$ means that σ' copies σ, and $A(s)$ denotes s together with its ancestors. Conditional on W the error originates at a single node O with law μ on the nodes extended by an idiosyncratic node ι that is an ancestor of no source, $\mu(\iota)$ being the share of errors no source inherits. A certificate on source s reports ok on W iff $O \in A(s)$, and ok on R always. For a purchased set T write $A_T = \bigcap_{s\in T} A(s)$ for the common ancestors of T, with $A_\varnothing$ the whole node set, and $c(T) = \mu(A_T)$; the value of T to one receiver is $V(T) = pL\,[1 - c(T)]$, and the marginal value of s given T is $\Delta_s(T) = V(T \cup \{s\}) - V(T)$. Let S be the set of all available sources.

Theorem 6 (Correlated roots)**.** *1. $F_A(T) = p\,\mu(A_T)$ and $F_R(T) = 0$: the all-ok probability is the mass of the common ancestors of the purchased sources.*

2. V is monotone and submodular, and $\Delta_s(T) = pL\,\mu(A_T \setminus A(s))$. It is zero iff $\mu(A_T \setminus A(s)) = 0$, in particular whenever s descends from every common ancestor of T: a copy of the copies adds nothing. It equals the whole residual $pL\,c(T)$ iff $\mu(A_T \cap A(s)) = 0$: a source whose provenance is disjoint from the common ancestors of T catches everything that is left.

3. (Floor.) $c(T) \geq c_\infty := \mu(A_S)$ for every purchasable T, with equality iff $\mu(A_T \setminus A_S) = 0$. If every available source descends from one node ρ with $\mu(\rho) = c_0$, then $F_A(T) \geq pc_0$ for every T: no number of certificates from one root lowers the floor.

4. (Against independence.) For two sources with marginal shares $c_s = \mu(A(s))$ and $c_{s'}$, $c(\{s, s'\}) \in [\max\{0, c_s + c_{s'} - 1\},\ \min\{c_s, c_{s'}\}]$, equal to the minimum when the ancestries are nested and to 0 when $A(s) \cap A(s')$ carries no mass. Shared ancestry makes a pair worth less than independent contamination with the same marginals would; disjoint ancestry makes it worth more.

Proof. (1) All certificates in T report ok on W iff $O \in A(s)$ for every $s \in T$, i.e. $O \in A_T$; on R all report ok, and accepting after all ok and rejecting after any flag is Bayes-optimal under (1) because $\mu(A_T) \leq 1$. (2) $V(T) = pL\,\mu\big(\bigcup_{s\in T} A(s)^c\big)$ is the measure of a union of sets, hence monotone and submodular, and $\Delta_s(T) = pL\,\mu\big(A(s)^c \cap A_T\big)$; both criteria read off this formula, the first because $A_T \subseteq A(s)$ whenever s descends from every node of A_T. (3) $A_T \supseteq A_S$ for $T \subseteq S$, and $\rho \in A_S$ when every source descends from ρ. (4) Fréchet bounds for the intersection of two events; nested ancestries give $A(s) \cap A(s') = A(s)$ and disjoint ones the empty set. □

Statement (4) corrects a natural intuition: "independent" certificates are not the ideal. Under a single origin, two sources whose ancestries share no mass jointly catch every inherited error, which independent contamination with the same marginals never achieves, and two registries copying one root jointly catch no more than one of them. The number of certificates measures nothing; the geometry of their ancestry does. Every statement of the theorem is checked on 300 random graphs and 2476 purchased sets; on these graphs the common-ancestor mass of a pair falls below the product of the marginals on 339 pairs and above it on 531.

Corollary 4 (One root, many copies)**.** *Let every available source be a child of one root ρ with $\mu(\rho) = c_0$, each source s carrying its own error mass $\mu(s) = a$. Then $c(\{s\}) = c_0 + a$ and $c(T) = c_0$ for $|T| \geq 2$: the second certificate catches exactly the errors that originated at the first source, and a third adds nothing. In the buyers' game at a common price K the pure-equilibrium number of certificates is 2 if $K < pLa$, 1 if $pLa < K < pL(1 - c_0 - a)$ and 0 if $K > pL(1 - c_0 - a)$; the social optimum uses the same thresholds with npL, so the bands $(pL(1 - c_0 - a),\ npL(1 - c_0 - a))$ for the first certificate and $(pLa,\ npLa)$ for the second are the wedge of Proposition 3, and the floor pc_0 is reached by two certificates and never lowered.*

Proof. $A(s) = \{\rho, s\}$, so $A_T = \{\rho\}$ for two or more distinct sources. The marginal values are $pL(1 - c_0 - a)$ for the first certificate, pLa for the second and 0 afterwards; only the count matters, and the argument of Theorem 5 applies with these increments. □

Proposition 2 (Equilibrium with correlated roots)**.** *In the buyers' game at posted prices P_s with public verdicts, if T is a pure-equilibrium outcome then $P_s \geq \Delta_s(T)$ for every $s \notin T$ and $P_s \leq \Delta_s(T \setminus \{s\})$ for every $s \in T$. With prices equal to real costs K_s, every source $s \notin T$ with $\Delta_s(T) < K_s < n\Delta_s(T)$ is efficient to add and bought by nobody: the band (v, nv) with $v = pL\,\mu(A_T \setminus A(s))$.*

Proof. Adding s changes a receiver's cost by $P_s - \Delta_s(T)$, dropping her own s by $\Delta_s(T \setminus \{s\}) - P_s$, and the social value of adding s is $n\Delta_s(T)$. □

Remark 2 (Bundles). The two conditions are necessary but not sufficient, because a receiver may replace several of her certificates by one. Let e be the parent of a and b, with $\mu(A(e)) = 0.1$ and $\mu(a) = \mu(b) = 0.3$, and let the prices be $P_a = P_b = 0.25$ and $P_e = 0.4$ in units of pL. Then $V(\{a, b\}) = V(\{e\})$: no single addition, drop or swap is profitable for the owner of a and b, but replacing both by e saves 0.1. A full characterization must consider every replacement of a receiver's own sub-bundle; the exchangeable case (Theorem 5) and the two-root case (Proposition 1) are the ones where single conditions suffice. On 150 random priced instances the necessary conditions hold at all 155 equilibrium outcomes, and 51 purchased sets satisfy them without being equilibria.

Aicumene:2026.048v1 [econ.TH] October 2026

6 Hidden source choice

We now remove observability. One provider sells a public verdict used by n receivers, or providers compete for the contract as in Section 4. After agreeing to provide the verdict for price P, the winning provider privately selects a source technology $t \in \{C, I\}$. Technologies, costs and losses are as in Section 4; the expected social residual loss after technology t is $npLc_t$. We abstract from who bears the sale price, which is a transfer, and suppose the contract is purchased and the public verdict distributed.

Timing: (i) a provider or competing firms post or accept a verification contract at price P, (ii) a buyer purchases the contract on the basis of equilibrium beliefs, (iii) the selected provider privately chooses t, (iv) verdicts are issued and all receivers act. Source provenance quality is not contractible unless an explicit audit is added. Buyers cannot infer the hidden source choice from the realized verdict of one claim; for the benchmark hidden-action argument they have no verifiable signal permitting payment to depend on t. This is a *moral-hazard* environment, not an adverse-selection ("lemons") theorem.

Theorem 7 (Hidden-action source degradation)**.** *If provider compensation does not depend on realized source choice, every profit-maximizing supplier chooses C after the sale. This remains true under any amount of ex-ante price competition among firms that face the same strictly positive independence cost premium ΔK. If*

$$0 < \Delta K < npL(c_C - c_I), \tag{12}$$

then this privately optimal source choice is strictly socially inefficient conditional on procurement; the aggregate resource-and-error welfare loss is exactly

$$npL(c_C - c_I) - \Delta K > 0. \tag{13}$$

Proof. After payment the supplier obtains $P - K_C$ from C and $P - K_I = P - K_C - \Delta K$ from I. Hence C is uniquely optimal at the hidden-action stage. The argument applies to the winner of any price competition because ex-ante rival prices do not enter the post-contract cost comparison. Aggregate real cost under C is $K_C + npLc_C$; under I it is $K_I + npLc_I$. The latter is lower if and only if (12) holds, by the positive difference (13). □

Scope. Existence of an equilibrium with a traded verification contract also requires willing buyers and a viable provider. For example, under common buyer funding, risk-neutral acceptance, and $P \in [K_C, pL(1-c_C))$, a buyer would purchase the expected cheap check rather than none. This assumption is not implicit in Theorem 7; the theorem characterizes *conditional service quality* and should not be read as a general equilibrium result for arbitrary procurement institutions. Firms with enforceable reputations, repeated contracting or external source audits need not degrade quality.

Theorem 8 (Audit threshold for genuine independence)**.** *Suppose a verifiable audit detects a shared-source choice C with probability $a \in [0, 1]$ and imposes a real contractual transfer or penalty $F > 0$ on the provider, while I never triggers the penalty. Then I is the unique post-contract supplier best response exactly when*

$$aF > \Delta K. \tag{14}$$

Under condition (12)*, the monitoring mechanism aligns the source-choice incentive with the cost-minimizing social technology. At equality the seller is indifferent.*

Proof. The expected seller profit under C becomes $P - K_C - aF$ and under I remains $P - K_I$. The latter exceeds the former if and only if $aF > \Delta K$. The social comparison was established in Theorem 7. □

The audit threshold is a contractual enforcement result, not a theorem that sanctions are socially costless. A full welfare analysis must include actual audit resources, distortionary enforcement costs, limited liability, detection errors, and long-term reputation.

7 The independence wedge: a synthesis

The six results share one arithmetic. In each environment a single decision-maker values source independence at some $v > 0$, the n decisions that rely on the public verdict value it at nv, and independence costs a premium δ paid once. Table 1 lists v, the band and the remedy.

Environment	private value v	band $v < \delta < nv$	remedy
One-shot procurement (App. A)	pcL	$pcL < d < npcL$	subsidy $s > d - pcL$
Sequential, private signals (Cor. 1)	$\frac{pc\lambda L}{pc\lambda + 1 - p}$	$v < \kappa_I < nv$	funded first purchase
Competitive supply, one certificate (Thm. 3)	$pL(c_C - c_I)$	$pL(c_C - c_I) < \Delta K < npL(c_C - c_I)$	collective financing
Multi-origin market, $(k{+}1)$st root (Thm. 5)	$pL(1-c)c^k$	$v_k < K < nv_k$ at every k	funded entry at each margin
Correlated roots, source s given T (Prop. 2)	$pL\,\mu(A_T \setminus A(s))$	$v < K_s < nv$	disjoint provenance
Hidden source choice (Thms. 7–8)	0	$0 < \Delta K < npL(c_C - c_I)$	audit $aF > \Delta K$

Table 1: The independence wedge in six environments. In each, independence is bought or supplied iff $\delta < v$ and is efficient iff $\delta < nv$; the band has width $(n-1)v$ and is empty iff $v = 0$, i.e. iff the shared source carries no common-cause error ($c = 0$, $c_C = c_I$, or $\mu(A_T \setminus A(s)) = 0$). In the sequential environment the values are conditional on receiver 1's flag signal; in the multi-origin market the band recurs at every diversity margin and shrinks geometrically.

Proposition 3 (Independence wedge)**.** *In each of the six environments, with the notation of Table 1, independence is provided in equilibrium iff $\delta < v$, it is efficient iff $\delta < nv$, and on the band $v < \delta < nv$ the welfare loss is $nv - \delta$. The primitive c enters only through v, and the number of relying decisions only through the factor n.*

Proof. Each row collects the two thresholds and the loss of the result named in it; in the multi-origin market the row applies at each k with $v_k = pL(1-c)c^k$ (Corollary 3). In the hidden-source environment the provider's private value of independence is zero because compensation cannot depend on it, so the band is $(0, nv)$, and an audit with sanction F detected with probability a raises the private value to aF (Theorem 8). □

What the six results add to the private provision of public goods [9] is the identity of the good. The under-provided good is the independence of evidence; its private value is governed by the common-source share c and vanishes with it; the decisions that rely on one verdict multiply its social value; and the information that would correct a common-source error can be in every receiver's hands (Theorem 2) and still go unused, because the public verdict settles each receiver's decision before her private information is consulted.

8 Public information, reliance, and the original contribution

Public signals can be valuable and still be highly dependent on one origin. Morris and Shin [4] investigate welfare effects of public information alongside private information and strategic complementarities; Svensson [5] cautions against treating transparency as generally welfare-reducing. Our sequential theorem does not reproduce their coordination externality. It addresses an information-identification constraint: when public actions and reported checks are measurable functions of a source that fails to distinguish two states, more observers cannot generate the missing distinction. Its relation to observational equivalence and classical

Aicumene:2026.048v1 [econ.TH] October 2026

Aicumene:2026.048v1 [econ.TH] October 2026

information cascades is direct, and novelty claims must be limited accordingly [10, 11]. What Theorem 2 adds to the cascade literature is the role of a public signal that is silent on one event: it lowers the residual prior to pc, so that the cascade starts before any private information has entered the public history, and it raises the precision a private signal needs by the factor $1/c$. What Theorem 3 adds to the moral-hazard reading of the supply side is that the under-provision of independence is first of all a public-good problem, present with fully observable quality and under price competition, with the same factor n as the demand-side wedge; Proposition 3 makes the common structure explicit. On the AI side, our experiments with cross-family LLM agents [3] find that such agents reach Nash equilibrium exactly when it reduces to a myopic best response over explicit payoffs; the sequential game of Theorem 2 is of that kind, since each receiver's decision is a myopic Bayes action on explicit losses, so its entrenchment prediction is directly testable with LLM receivers, whereas the one-shot procurement wedge requires anticipating free riding and is not.

The new price-quality game provides a complete equilibrium for a specified procurement institution, while the earlier hidden-action comparison studies a different information friction. The seller-side model moves from fixed alternative verification services to the incentive of an actual provider to supply independently sourced evidence. The zero-audit degradation result is a simple instance of unobservable effort layered on the public-good wedge, not a broad theorem that all competitive markets select correlated signals. Its economic significance comes from connecting this incentive problem to a common-source information floor and to public reliance by n recipients: the social value of independence rises with the number who trust the public verdict, while the provider's private cost difference need not.

9 Measurement proposal: legal-source verification

A feasible empirical design, motivated by the companion *Verifiable AI for Law* architecture [2], would randomly assign controlled claim instances with a source-level truth label to (i) verifiers using the proposer's source snapshot and (ii) verifiers using independently curated authority snapshots. The experimental unit should be a documented claim–source pair, not a count of agreeable model outputs. Before estimating c, independently estimate each verifier's false-alarm and false-negative rates and compare the likelihood of inherited errors conditional on a wrong claim. Public verdict propagation can then be varied across simulated legal teams; costs should include evidence acquisition, review time and provenance validation. No real legal-case results are reported in this paper.

10 Limitations and next theory steps

Three assumptions do substantial work: the indistinguishability of correct claims and common-source errors, the binary symmetric private signals of the sequential game, and the non-excludability of the public verdict. Relaxing any may change the predicted failure. With heterogeneous or continuous private signals the cascade of Theorem 2 need not start at the first receiver, and the entrenchment bound becomes a statement about the distribution of the first informative action; with an excludable verdict the supply-side wedge closes and only the hidden-action layer remains. The one-certificate market theorem characterizes pure equilibria under a single-certificate registrar with priority funding and says nothing about mixed strategies or other funding equilibria; alternative allocation and joint-funding institutions remain open mechanism-design problems. The correlated-roots theorem keeps one origin per error; errors with several independent origins combine the product law and the ancestry law and are left open, as is the full equilibrium characterization with bundles (Remark 2).

The next nontrivial extensions are: the sequential game with heterogeneous signal precisions and with the shared check itself bought along the sequence; providers investing in a continuous provenance quality $c(z)$ with imperfect auditing and heterogeneous costs; equilibrium selection and welfare under multi-buyer financing of the public verdict; and evidence aggregation over partially overlapping source DAGs. A further extension replaces equilibrium play by learning agents; by the criterion of [3], LLM receivers should reproduce the cascade of Theorem 2, and the protocol of that paper applies directly. These extensions are open problems, not solved theorems of this note.

Aicumene:2026.048v1 [econ.TH] October 2026

11 Conclusion

The number of verifications is a poor substitute for independence of their informational origins. A shared-source mistake survives arbitrary public sequential reliance even when every receiver holds independent private information, because the public verdict starts the cascade before that information is used; every pure equilibrium of the certificate market buys the copied source whenever independence costs more than one receiver's value, the same wedge recurs at every margin of source diversity, and a provider facing an unobservable post-contract choice selects the cheaper copied source. The welfare case for independence scales with the number of receivers, while the post-contract supplier cost advantage of copying does not. We establish an observable-quality competitive posted-price equilibrium under a single-certificate priority registry. The unrestricted procurement extension now provides exact source-diversity entry thresholds and a restricted within-origin Bertrand implementation; it does not claim a fully unrestricted price–provenance supply equilibrium. The proved claims are conditional, elementary, and designed as a tractable foundation for the richer equilibrium theory still required for a strong economics-journal submission. In every environment the private value of independence is one n-th of its social value, so the band of under-provision widens with the number of decisions that rely on one verdict and closes only when the common source carries no error.

A Prior benchmark: procurement wedge

For reference, in a one-shot game with $n \geq 2$ identical receivers choosing among no check, a public shared check at price κ, or an independent check at price $\kappa + d$, own error $q = 0$ and (1), the private shared check is preferred to no check when $\kappa < pL(1-c)$. In the interior $pcL < d < npcL$ an individual check-buyer prefers the shared check to upgrading, but independent verification minimizes aggregate costs. This follows by comparing $\kappa + pcL$ with $\kappa + d$ privately and $\kappa + npcL$ with $\kappa + d$ collectively [1]. It is included for context, not counted as a new contribution.

B Imperfect signal robustness: a sufficient bound

The previous working paper gives a constructive continuity bound for binary own-error q and a local identity for the value of a second shared check. With $A = (1-p)\ell - pL > 0$, within the neighborhood where the Bayes rules are one-flag rejection for one shared check and two-flag rejection for two shared checks, the reduction in Bayes risk equals $Aq(1-q)$. This is a local result, not the first instability threshold of the full game. With $m = \min\{\kappa, d - pcL, pL(1-c) - \kappa\}$, $\Delta = \min\{npcL - d, npL - \kappa - d\}$, and $M = \max\{L, \ell\}$, a conservative radius for preserving the procurement-wedge equilibrium comparisons is

$$\overline{q} = \min\left\{\frac{1}{2}, \frac{m}{2nM}, \frac{\min\{\Delta, \kappa\}}{2n^2 M}\right\}. \tag{15}$$

A coupling argument bounds individual Bayes-risk shifts by nqM and aggregate shifts by $n^2 qM$. Full statements and numerical checks belong to the earlier working versions rather than the main economic mechanism here.

C Computational reproducibility

The file `market_equilibrium.py` exhaustively checks profitable firm deviations on a finite price grid and compares costs on a parameter grid; this complements, rather than proves, Theorem 3. The file `seq_model.py` computes the equilibrium path of the sequential game of Theorem 2 exactly, for every reachable public history and signal, on 2484 parameter points with up to $n = 8$ receivers, and confirms both cases of the theorem and the value v of the corollary; it draws Figure 1. The file `market_equilibrium.py` checks Theorem 3 on a price grid: at the all-(C, K_C) profile no firm has a profitable deviation for $n \in \{2, 3, 4, 6, 10\}$, and, enumerating all 14884 pure profiles of two firms (price step 0.005) and all 238328 profiles of three firms (step 0.01), every profile without a strictly profitable unilateral deviation implements C, bought by receiver 1, at a price within one grid step of K_C, the grid analogue of undercutting (6 and 3213 such profiles, differing only in the losing offers). The files `verify_multi.py` and `multi_origin.py` enumerate the buyers' game of Theorem 5 for

$n \leq 3$ and up to six roots, and `multi_check.py` confirms Proposition 1 on 248 parameter points of the two-root game, the count formula $k^{NE} = \min\{M, \lceil \log(K/B)/\log c \rceil\}$ with $B = pL(1-c)$ on 72 points, and produces the numbers of the illustration. The file `dag_check.py` checks Theorem 6 on 300 random source graphs (2476 purchased sets: origin-by-origin computation of the all-`ok` probability, floor, marginal-value formula, zero- and full-gain criteria, submodularity), Corollary 4 on 96 cases by enumeration of the buyers' game, the example of Remark 2, and Proposition 2 on 150 priced instances. The file `model_v9.py` checks the arithmetic of Theorems 7 and 8 and writes `results.csv` and the figure of Section 4. The plots describe illustrative parameters and should not be cited as empirical verification of real-market behavior. The code deliberately distinguishes a mathematical identity check from a validation of a model's external assumptions.

D Code Availability

The source code and implementation of the legal verifiability compiler, including the rule/judgment firewall, are available in the Aicumene GitHub repository: `github.com/aicumene/verification-games`.

The repository contains the implementation and supporting materials for the Verifiable AI for Law research project.

E Scope and limitations

All legal thresholds, effective dates, and citations must be independently verified by qualified counsel before authoritative use. The compiler only decides what can be expressed as reliable executable predicates over validated facts; it cannot autonomously resolve open-textured legal standards or ensure an encoded rule is legally correct. Claims about self-improvement are epoch-local. The system is presented for *research and legal decision support, not autonomous legal advice.*

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] Bernadotte, A. and Menshikov, I. (2026). Verifiable AI for law: a legal verifiability compiler with a rule/judgment firewall. Aicumene Preprint. `https://research.aicumene.com/papers/verifiable-law/`, doi:10.13140/RG.2.2.18174.75841.

[3] 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.

[4] Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review* 92(5), 1521–1534. doi:10.1257/000282802762024610.

[5] Svensson, L. E. O. (2006). Social value of public information: Comment: Morris and Shin (2002) is actually pro-transparency, not con. *American Economic Review* 96(1), 448–452. doi:10.1257/000282806776157650.

[6] Morris, S., Shin, H. S. and Tong, H. (2006). Social value of public information: Reply. *American Economic Review* 96(1), 453–455. doi:10.1257/000282806776157597.

[7] 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.

[8] 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.

Aicumene:2026.048v1 [econ.TH] October 2026

[9] 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.

[10] 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.

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

[12] Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84(3), 488–500. doi:10.2307/1879431.

Aicumene:2026.048v1 [econ.TH] October 2026