The Game of Peers: Toward a Credential Algebra for Human-AI Actor Networks

Yaroslav Ryabov

Abstract—Current authorization frameworks assume humans delegate authority downward to AI agents and commit to monotonic restriction: adding an actor can only narrow permissions. We introduce a *credential algebra* in which human and AI actors are structural peers, each carrying a credential set combined by set operators (intersection, union, and two joins) on the edges of a directed actor graph; classical delegation is recovered as intersection on a rooted tree. Our results characterize when the monotonic worldview survives and what replaces it. A dichotomy theorem shows that coalition value is guaranteed monotone if and only if every edge passes its subordinate endpoint unfiltered—no restriction on topology can substitute—on trees and, unchanged, on arbitrary directed graphs under a least-solution semantics; for a deployed assignment of credentials, a linear-time *protection* criterion decides whether any credential can ever be lost. An exact criterion identifies *destructive intermediaries*: an actor's marginal contribution is negative precisely when the credentials that only its filtered children hold, and that its filters destroy, outnumber its novel contribution; leaf joiners are provably harmless, so the algebraic cost of oversight lies in its placement, not its presence. Deciding whether an actor can ever be destructive is NP-complete already on depth-one trees with two operators, yet fixed-parameter tractable in the number of credentials on trees, and hard again on cyclic graphs even with a single credential; evaluating a given coalition is linear-time, and Shapley values admit polynomial-time sampling approximation. A strategic layer classifies when pure-strategy disclosure equilibria exist. Every question answered above is unposable under monotonic restriction; the paper's contribution is thus twofold: making these questions expressible, and answering them.

Index Terms—Access control, delegation, trust management, authorization policies, multi-agent systems, human-AI collaboration, cooperative game theory, computational complexity.

1 Introduction

Authorization for autonomous AI agents is being standardized now. The protocols that extend OAuth 2.0 and OpenID Connect to agents [1], the certificate-chain tradition of SPKI/SDSI [2], and the cryptographic scope narrowing of authenticated workflows [3] share one design commitment: *attenuation only*. A human principal is the root of trust; an agent's authority is a subset of its delegator's; adding an actor to a chain can only narrow what the chain is authorized to do. In the classical threat model this is exactly right: amplification of authority along a delegation chain *is* the attack, and ruling it out by construction was sound engineering.

What has changed is the threat model. Agents now acquire credentials not derived from any human (a transaction-monitoring certification, a sanctions-screening attestation), initiate workflows without a human prompt, and collaborate across organizations that share no authority chain. In such networks a second failure mode appears alongside over-authorization: a coalition of actors can be authorized to do too *little*. A compliance intermediary placed on a path "to be safe" can silently remove a credential that no other actor holds and that the task requires—while every local policy is obeyed and no component misbehaves. Monotonic restriction protects the *safety* of authorization (nothing exceeds its delegator) at the price of its *availability*: the task's authorization can vanish, and, because the formalism's language cannot express the loss, the failure is invisible to the very model meant to reason about it. This is a failure of authorization *sufficiency*, not of confinement, and it is a dependability problem as much as a security one. Independent experiments already observe the symptom: adding actors or human oversight can degrade collective outcomes [4], [5]—an observed phenomenon still awaiting its formal question.

Why existing authorization formalisms cannot express it: Three structural absences in the formalisms available today do the blocking. Monotone frameworks—trust management, certificate chains, and cooperative games with permission structures [6]—make harm *inexpressible*: adding an actor narrows permissions by axiom, so "can this actor's presence hurt?" is a category error. Policy composition algebras [7], [8], which do admit non-monotone composition, make blame *unattributable*: an expression has operators but no owners. And consent-level games make damage *undiagnosable*: edges carry permission bits, not the content that could be lost. No prior formalism assembles actors that own content, operators that can destroy it, and coalitional counterfactuals that vary it—the three ingredients every question below requires.

Peers, not delegates: We propose a framework in which all actors—human and artificial—are *structural peers* with independent credential sets. No actor is *a priori* the root, delegate, or subordinate of another. Effective authorization is computed by set-algebraic operators on the edges of a directed actor graph: intersection, union, and two asymmetric joins that pass a subordinate's credentials through a relevance filter. These four capture a broad and practically

• *Y. Ryabov is an independent researcher in Rockville, MD, USA. E-mail: yaroslav.ryabov@gmail.com.*

important class of authorization patterns (we claim no exhaustiveness), and classical delegation is recovered exactly as intersection on a rooted tree. Because credentials are sets and edges transform them, the framework generates, for every subset of actors, a characteristic function that prior models could only restrict from outside; the cooperative-game machinery—marginal contributions, Shapley attribution—then applies to authorization itself.

Contributions: The paper's contribution is twofold: it makes a class of questions about authorization expressible, and it answers them, up to and including proofs (see the supplemental material) of which among them are computationally answerable at all.

(i) *A credential algebra with a coalition semantics* (Section 3) on arbitrary directed graphs, cycles included. Effective credentials are the *least solution* of a system of propagation equations, one per actor: the smallest credential assignment consistent with every edge, containing exactly what can be derived from the query's origins along the edges and nothing more. On trees and acyclic graphs this is an ordinary recursion; in general it is the unique *grounded* solution—the same semantic choice that Datalog-based trust-management languages make [9].

(ii) *Exact certification of the monotone worldview*: dichotomy theorems (Theorems 4.1 and 4.3) showing that attenuation-only guarantees hold if and only if every edge passes its subordinate endpoint unfiltered—so no restriction on topology can substitute for the operator condition—and its counterpart for a deployed assignment, a linear-time protection criterion deciding whether any credential can ever be lost (Proposition 4.1).

(iii) *An audit metric and a placement rule*: a closed-form criterion for destructive intermediaries (Theorem 4.2) and the harmlessness of leaf and sink joiners (Propositions 4.2, 4.4), so that the algebraic cost of human oversight lies in its placement, not its presence.

(iv) *A complexity boundary in the tradition of Harrison–Ruzzo–Ullman*: auditing a configuration is linear-time; deciding whether an actor can ever be destructive is NP-complete already on depth-one trees with two operators, so certifying that it never is, coNP-complete (Theorem 6.1); counting destructive coalitions is #P-complete; all three are nonetheless fixed-parameter tractable in the number of credentials on trees (Proposition 6.3), while cycles are an independent source of hardness—coNP-complete with a single credential (Proposition 4.3); Shapley attribution is approximable by sampling; and minimum-cost coalition synthesis is weighted set cover in the preserving regime and loses submodular structure once filters appear (Section 6).

(v) *Disclosure incentives*: a classification of when pure-strategy disclosure equilibria exist (Section 5), including the reading of the no-disclosure deadlock on intersection edges as a coordination failure.

Each of these answers attaches to a decision that architects placing humans in agent loops, regulators attributing responsibility, and auditors certifying configurations are already making without a formal basis. A six-actor cross-border stress test (Section 4.2; Example 4 in the supplemental material, where three further worked examples precede it) instantiates the results; its numbers follow from the criterion by substitution. Proofs, the explicit computations for all examples, and the full discussion of modeling assumptions are in the supplemental material.

2 Related Work

Trust management and policy algebras. Treating authorization algebraically originates with PolicyMaker and KeyNote [10] and with SPKI/SDSI, whose certificate chains compose permission sets by intersection [2]—the operation our framework recovers as classical delegation (Proposition 3.3). Li and Mitchell [11] gave SPKI/SDSI a logical semantics and noted that its tuple reduction is semantically incomplete in part because it does not handle *union* of authorization tags; the RT framework [9] fixed credential semantics as the least model of a Datalog translation, the same grounded semantics we adopt on cyclic graphs (Section 4.3). A separate line composes *policies* rather than certificates: Bonatti, De Capitani di Vimercati, and Samarati [7] introduced an algebra of access control policies with union, intersection, and difference; Wijesekera and Jajodia [8] gave a propositional policy algebra; and XACML's combining algorithms embody the same pluralism industrially [12]. These works established, two decades before the current agent protocols recommitted to monotonic restriction, that non-monotone composition is a recognized need. We concede the closest kinship here: for a *fixed* coalition, propagation on a tree evaluates a nested operator expression, inter-translatable with a policy expression (the operator inventories differ—theirs include difference and scoping, ours parameterized joins). The two representations differ only in where the operators live: a policy expression puts atomic policies at the leaves and operators at the internal nodes, whereas an actor graph puts the atoms (credential sets) at the actors and the operators on the edges (the converse translation holds only for expressions in which each operator has an atomic operand). What has no counterpart is the *coalition-indexed family* of expressions one actor graph generates—varying the set of actors reassembles the expression tree, an operation a policy expression does not possess: each subset of actors induces its own forest, re-rooting components and reverting absent actors' subtrees to intrinsic contributions (Definition 3.11). That family—a characteristic function over 2^A—with actors who *own* the leaves is what makes destructive intermediaries, Shapley attribution, and the complexity of destructiveness definable at all.

Cooperative games with restricted cooperation. Myerson's communication games [13] restrict cooperation to connected coalitions. Closest to us are games with *permission structures* [6]: players of a transferable-utility game sit in a directed hierarchy and need permission from all (conjunctive) or at least one (disjunctive) predecessor. The parallel is real and so is the difference. There the digraph carries *consent* restricting an *exogenous* game, and the restriction provably preserves monotonicity—adding a superior can only unlock subordinates. Here the digraph carries *content* and *generates* the game; edges transform credential sets. The destructive intermediary is therefore invisible in the permission-

structure class, and our dichotomy characterizes exactly when content-level composition collapses to the monotone behavior consent-level restriction always exhibits. The conjunctive/disjunctive split is intersection versus union at the level of *who must consent*; our operators act at the level of *what passes through*.

AI agent delegation and identity. South et al. [1] extend OAuth 2.0 and OpenID Connect to agents, and Authenticated Workflows [3] enforce cryptographic scope narrowing—delegated scope equals the intersection of what the delegator grants and the delegate possesses—both firmly within attenuation-only delegation. Identity-layer efforts based on decentralized identifiers and verifiable credentials [14] address storage and verification, not composition. Shapley-based pricing of agents' *operational* contributions is a recent theme; ours attribute *authorization potential*, and negative attributions arise structurally from filtering edges. The multi-agent trust literature models trust as a scalar to be estimated and updated [15]; we model authorization as a set subject to algebra, with scores entering only as parameters of the strategic layer.

The gap. Traditional access control treats the human as an unqualified root of trust who merely *grants* subsets of its authority [16]. In cross-organizational peer collaboration the failure mode is no longer only that a coalition can do too much, but that it can do too little—capability quietly destroyed by composition—and a formalism that cannot express the second failure cannot price it. A human connected through a filtering operator can be *algebraically destructive*: the operator on its edge removes more effective credentials than the human's own set contributes, so the coalition is authorized to do strictly more without it. Making that sentence formal, and then deciding it, is the program of this paper.

3 The Credential Algebra

3.1 Actors, Credentials, and Operators

Definition 3.1 (Actor Universe)**.** *Let $\mathcal{A}$ be a finite set of* actors*. Each actor $a \in \mathcal{A}$ is either human ($a \in \mathcal{A}_H$) or artificial ($a \in \mathcal{A}_M$), with $\mathcal{A} = \mathcal{A}_H \cup \mathcal{A}_M$ and $\mathcal{A}_H \cap \mathcal{A}_M = \emptyset$. We impose no structural distinction between the two classes: both are first-class participants in the framework.*

Definition 3.2 (Credential Universe)**.** *Let $\mathcal{C}$ be a finite set of* atomic credentials*. Each credential $c \in \mathcal{C}$ represents an indivisible unit of authorization, certification, capability, or trust—for example, a permission, a regulatory accreditation, a reputation attestation, or a security clearance.*

Definition 3.3 (Intrinsic Credential Set)**.** *Each actor $a \in \mathcal{A}$ carries an* intrinsic credential set *$\sigma(a) \subseteq \mathcal{C}$: the credentials that a holds in its own right, independently of any collaboration. The function $\sigma : \mathcal{A} \to \mathcal{P}(\mathcal{C})$ is the* credential assignment*. The qualifier* intrinsic *distinguishes $\sigma(a)$ from two derived notions introduced later: the* effective *credential set $\hat{\sigma}(a)$, which results from combining credentials along the actor graph (Definition 3.9), and the* disclosed *credential set $\delta_{a\to b} \subseteq \sigma(a)$, which is the subset an actor strategically chooses to reveal (Section 5.1).*

The partition of actors into humans and machines is metadata: no definition below treats the two differently, and the theory's statements are actor-type-blind.

Definition 3.4 (Actor Graph)**.** *The* actor graph *is a directed graph $\mathcal{G} = (\mathcal{A}, E)$ where $E \subseteq \mathcal{A} \times \mathcal{A}$ is the set of directed edges. An edge $(a, b) \in E$ represents a collaboration relationship in which actor a's credentials interact with actor b's credentials. The direction encodes the flow of a task or request.*

$\mathcal{G}$ is not assumed acyclic (Section 4.3 treats cycles), and symmetric interactions carry both (a, b) and (b, a), possibly with different operators.

Definition 3.5 (Credential Operator)**.** *A* credential operator *is a function $\otimes : \mathcal{P}(\mathcal{C}) \times \mathcal{P}(\mathcal{C}) \to \mathcal{P}(\mathcal{C})$ that takes the credential sets of two actors and produces a combined credential set. Operators may be drawn from a* parameterized family*: for a parameter set $R \subseteq \mathcal{C}$, we write $\otimes_R$ for the member of the family indexed by R; each such $\otimes_R$ is itself a binary operator of the above type. (The join operators of Definition 3.6 form such families, indexed by a relevance filter R; intersection and union are unparameterized.) Each edge $(a, b) \in E$ is equipped with a fully instantiated binary operator $\otimes_{a,b}$—including, where applicable, a fixed value of its parameter R—giving rise to an* operator assignment *$\Omega : E \to (\mathcal{P}(\mathcal{C}) \times \mathcal{P}(\mathcal{C}) \to \mathcal{P}(\mathcal{C}))$.*

Definition 3.6 (Fundamental Operators)**.** *The four fundamental operators are intersection $\cap$, union $\cup$, and, for a* relevance filter *$R \subseteq \mathcal{C}$ fixed on the edge, the left join $X \ltimes_R Y := X \cup (Y \cap R)$ and the right join $X \rtimes_R Y := Y \cup (X \cap R)$ (Table 1). When R is clear we write $\ltimes$, $\rtimes$.*

Proposition 3.1 (Properties of the Fundamental Operators)**.** *All four operators are idempotent, monotone in both arguments, and bounded ($\otimes(X, Y) \subseteq X \cup Y$); intersection and union are commutative and associative; the joins are neither (Table 1).*

Non-associativity of the joins is not a defect but a feature of relevance filtering: which filter is applied first changes what the next filter sees, so composition along a path is inherently *ordered*—a property the propagation semantics below encodes.

Definition 3.7 (Pass Filter)**.** *The* pass filter *of a credential operator $\otimes$ is the largest set $\pi(\otimes) \subseteq \mathcal{C}$ such that $\otimes(X, Y) \supseteq Y \cap \pi(\otimes)$ for all $X, Y \subseteq \mathcal{C}$: the part of the child's set that the edge passes whatever the parent holds. An edge is* preserving *if its operator has $\pi = \mathcal{C}$ and* filtering *otherwise.*

The four operators are the instances that matter in practice, but the structural and complexity results below never use their specific form—only that each passes the child's credentials through a fixed filter and the parent's through a filter that may depend on the child. We isolate that class.

Definition 3.8 (Pass-Filtered Operators)**.** *A credential operator $\otimes$ is* pass-filtered *if there are a set $\pi \subseteq \mathcal{C}$ (the* child filter*) and a monotone map $\alpha : \mathcal{P}(\mathcal{C}) \to \mathcal{P}(\mathcal{C})$ (the* parent filter*) with $\alpha(Y) \supseteq Y \setminus \pi$ for all Y, such that*

$$\otimes(X, Y) \;=\; \big(X \cap \alpha(Y)\big) \cup (Y \cap \pi) \qquad \text{for all } X, Y \subseteq \mathcal{C}.$$

We write $\mathcal{P}$ for the class of pass-filtered operators. The four fundamental operators are pass-filtered: $\cap$ has $\alpha(Y) = Y$, $\pi = \emptyset$; $\cup$ has $\alpha \equiv \mathcal{C}$, $\pi = \mathcal{C}$; $\ltimes_R$ has $\alpha \equiv \mathcal{C}$, $\pi = R$; and $\rtimes_R$ has $\alpha \equiv R$, $\pi = \mathcal{C}$.

Lemma 3.1 (Properties of Pass-Filtered Operators)**.** *Every $\otimes \in \mathcal{P}$ with filters (α, π) is monotone in both arguments,*

TABLE 1
The four fundamental operators on an edge (p, a), applied as $\otimes(X, Y)$ with X the parent's effective set and Y the child's credential set. π is the pass filter (Def. 3.7). All four are idempotent, monotone in both arguments, and bounded ($\otimes(X, Y) \subseteq X \cup Y$).

Operator	$\otimes(X, Y)$	$\pi(\otimes)$	Commutative	Associative	Authorization reading
Intersection $\cap$	$X \cap Y$	$\emptyset$	yes	yes	both must agree (compliance-critical)
Union $\cup$	$X \cup Y$	$\mathcal{C}$	yes	yes	pool capabilities (peer collaboration)
Left join $\ltimes_R$	$X \cup (Y \cap R)$	R	no	no	parent primary; the child's credentials pass only through the relevance filter R
Right join $\rtimes_R$	$Y \cup (X \cap R)$	$\mathcal{C}$	no	no	child primary; the parent's credentials pass only through R

bounded ($\otimes(X, Y) \subseteq X \cup Y$), and union-distributive in its first argument; its pass filter (Definition 3.7) is $\pi(\otimes) = \pi$; and for all X, Y,

$$Y \cap \pi \;\subseteq\; \otimes(X, Y) \;\subseteq\; X \cup (Y \cap \pi),$$
$$\otimes(X, Y) \cap Y \;=\; Y \cap (X \cup \pi).$$

Every result of Sections 4 and 6 stated for the fundamental operators holds verbatim for $\mathcal{P}$; the strategic results of Section 5 additionally use locality (membership of a credential in $\otimes(X, Y)$ depends only on its membership in X and Y), which the four operators have and a general parent filter need not.

3.2 Propagation and Coalition Effective Sets

Definition 3.9 (Effective Credential Set Along a Path)**.** *Let $p = (a_0, a_1, \ldots, a_n)$ be a directed path in $\mathcal{G}$. The* effective credential set *along p is defined recursively:*

$$\hat{\sigma}_p(a_0) \coloneqq \sigma(a_0) \tag{1}$$
$$\hat{\sigma}_p(a_k) \coloneqq \otimes_{a_{k-1}, a_k}(\hat{\sigma}_p(a_{k-1}),\ \sigma(a_k)) \quad \text{for } k = 1, \ldots, n \tag{2}$$

The terminal value $\hat{\sigma}_p(a_n)$ is the effective credential set at the end of the path.

Definition 3.10 (Credential Algebra)**.** *A* credential algebra *is a tuple $(\mathcal{C}, \mathcal{G}, \{\otimes_e\}_{e \in E})$: a credential universe, an actor graph $\mathcal{G} = (\mathcal{A}, E)$ (Definition 3.4), and a pass-filtered operator $\otimes_e \in \mathcal{P}$ (Definition 3.8) on every edge. Boundedness (no credential creation) and monotonicity (more input credentials, no fewer output credentials), which every pass-filtered operator satisfies (Lemma 3.1), are the minimal requirements for meaningful authorization semantics; the uniform child filter of $\mathcal{P}$ is what makes the dichotomy and the destructive-intermediary criterion possible.*

Much of what follows studies how effective credentials behave as the *set of participating actors* varies. We therefore fix, at the algebraic level, the effective set of an arbitrary coalition—on an arbitrary directed graph, as the least solution of a system of propagation equations, of which the recursion along a tree is the special case used in the examples.

Definition 3.11 (Coalition Effective Set)**.** *Let $\mathcal{G}$ be a directed graph without self-loops and $O \subseteq \mathcal{A}$ a set of* origins*: the actors at which a query starts (unless stated otherwise, $O = \emptyset$). For $S \subseteq \mathcal{A}$ let $F_S = (S, E_S)$ be the subgraph induced by S; the* parents *of a in F_S are $\mathrm{par}_S(a) \coloneqq \{p \in S : (p, a) \in E\}$, and the* sources *of F_S are the actors without parents, $\mathrm{src}(F_S) \coloneqq \{a \in S : \mathrm{par}_S(a) = \emptyset\}$. The* origins of the coalition *are*

$$O_S \coloneqq (O \cap S) \cup \mathrm{src}(F_S) :$$

designated origins in S, together with the actors that S cuts off from every parent and which therefore revert to their intrinsic sets. (On a rooted tree the root has no parent, so it is a source of every coalition containing it, and no origin need be designated.)

A credential assignment *on S is a map $E : S \to \mathcal{P}(\mathcal{C})$, one candidate effective set per actor; assignments are ordered pointwise, $E \leq E'$ if $E(a) \subseteq E'(a)$ for all a, so that $\mathcal{P}(\mathcal{C})^S$ is a finite complete lattice whose least element $\bot$ assigns $\emptyset$ to every actor. Write $\sigma_{O_S}(a) \coloneqq \sigma(a)$ if $a \in O_S$ and $\emptyset$ otherwise. The* propagation map *$\Theta_S : \mathcal{P}(\mathcal{C})^S \to \mathcal{P}(\mathcal{C})^S$ sends an assignment E to the assignment*

$$\Theta_S(E)(a) \coloneqq \sigma_{O_S}(a) \;\cup \bigcup_{p \in \mathrm{par}_S(a)} \otimes_{(p,a)}(E(p),\, \sigma(a))\,, \tag{3}$$

in which every actor receives its intrinsic set if it is an origin of the coalition, plus whatever each parent's current set passes through the edge. An assignment E is a solution *if $\Theta_S(E) = E$, that is, a fixed point of Θ_S. Under full disclosure, the* effective credential sets within *S form the* least *solution—the solution below every other in the pointwise order, which exists by Proposition 3.2—denoted $\hat{\sigma}_S \coloneqq \mathrm{lfp}\,\Theta_S$, so that $\hat{\sigma}_S(a)$ is actor a's effective set; the* coalition effective set *is $\hat{\sigma}_S \coloneqq \bigcup_{a \in S} \hat{\sigma}_S(a)$. On a rooted tree F_S is a forest, every actor has at most one parent, and the least solution is the unique one, given by the recursion*

$$\hat{\sigma}_S(a) = \begin{cases} \sigma(a), & a \text{ a component root of } F_S, \\ \otimes_{(p,a)}(\hat{\sigma}_S(p),\, \sigma(a))\,, & p \text{ the parent of } a \text{ in } F_S; \end{cases} \tag{4}$$

in particular, an actor isolated in F_S contributes exactly its intrinsic set, whereas a member whose parent is present contributes only what its incoming edge passes: the coalition set is a union of effective*, not intrinsic, sets (under intersection a delegate exercises only what it shares with its delegator, Proposition 3.3). Equation* (3) *aggregates parents* disjunctively*; on trees nothing is aggregated, and the conjunctive alternative is treated in the supplemental material.*

Definition 3.12 (Coalition Value)**.** *A* valuation *is a function $W : \mathcal{P}(\mathcal{C}) \to \mathbb{R}_{\geq 0}$ that is monotone with respect to inclusion ($X \subseteq Y$ implies $W(X) \leq W(Y)$). Throughout the examples W is cardinality, $W(X) = |X|$. Under full disclosure the* value *of a coalition S is, in* shared-outcome *evaluation,*

$$v(S) \coloneqq W(\hat{\sigma}_S),$$

the valuation of everything the members of S can jointly assemble; in at-target *evaluation, $v(S) \coloneqq W(\hat{\sigma}_S(t))$ for a designated target actor t, with $v(S) = W(\emptyset)$ when $t \notin S$. The structural results of Section 4 are stated for shared-outcome evaluation; the supplemental material records what carries over to a target. v measures what a coalition* can *do, not whether it* may *act: a mandated overseer or attestation is a requirement $Q \subseteq \hat{\sigma}_S$ on the coalition (Definition 6.2), a constraint, not value.*

Zero disclosure costs make full disclosure optimal, so v as just defined is then the characteristic function of the coalition game of Section 5; with positive costs, the strategic layer operates on top of this baseline, with disclosed sets δ replacing σ in (3) (Definition 5.4); a disclosure profile is itself a credential assignment, so every result of Section 4 applies to it verbatim.

Proposition 3.2 (Least Solution and Evaluation)**.** *Θ_S is monotone, so its solutions form a nonempty complete lattice [17]; the least solution is the limit of $\Theta_S^k(\bot)$ from the empty assignment, stationary within $|S| \cdot |\mathcal{C}|$ steps of cost $O(|E_S| \cdot |\mathcal{C}|)$ each, so $\hat{\sigma}_S$, $v(S)$, and marginal contributions are polynomial-time computable on any directed graph. If F_S is acyclic the solution is unique and computed in one topological pass in $O((|S| + |E_S|)|\mathcal{C}|)$ time.*

The least solution is the union of the values of walks ending at each actor, and a solution is the least one if and only if it is *grounded*—each credential has such a derivation (supplemental material, Section S2).

Remark 3.1 (Why the least solution)**.** *On trees and acyclic graphs nothing is chosen. On cyclic graphs the equations can have many solutions: on the two-cycle $a \to b$ (intersection), $b \to a$ (union) with $O = \{a\}$, every $E(a) = \sigma(a) \cup T$, $T \subseteq \sigma(b) \setminus \sigma(a)$, is one. The least solution is the closure of the propagation process from the state in which only origins hold their sets; it is the only solution free of credentials supported solely by a cycle of mutual assumption (supplemental material), which is why Datalog-based trust management adopts least-model semantics [9]; and it is the most conservative, a default-deny reading. The greatest solution degenerates, and iterating from "every actor holds its intrinsic set" can oscillate with period two.*

Fixed points and dynamics: Solutions of (3) are the stationary states of propagation, not its dynamics. Deriving from the origins, as Datalog engines do, is the iteration from $\bot$ and terminates within $|S| \cdot |\mathcal{C}|$ rounds on any graph; iteration from the intrinsic state converges to the least solution when every actor's incoming pass filters cover $\mathcal{C}$, and may otherwise cycle (supplemental material; Section 7.3).

Proposition 3.3 (Delegation as Intersection on a Tree)**.** *Let $\mathcal{G}$ be a directed tree rooted at a human actor $h \in \mathcal{A}_H$, with all edges equipped with the intersection operator $\cap$. Write $a_0 = h$ and let $h = a_0 \to a_1 \to \cdots \to a_n$ be the path to an actor a_n at depth n. Then:*

$$\hat{\sigma}(a_n) = \sigma(h) \cap \sigma(a_1) \cap \cdots \cap \sigma(a_n) \subseteq \sigma(h) \tag{5}$$

That is, the effective credential set at any point in the tree is a subset of the root's credentials. This recovers the classical property that delegation can only narrow permissions.

Scope and assumptions: The results are stated for the exact graph class they hold on (Section 4.3). The worked examples adopt binary, atomic credentials, keep actor scores as a separate structure entering only the strategic layer, use the four fundamental operators (the theory holds for $\mathcal{P}$), and assume sequential disclosure where a protocol is needed; the supplemental material discusses and relaxes each. Finally, the algebra prices failures of authorization *availability*; it does not replace the analysis of confinement, and the two are complementary: the dichotomy says when attenuation-only guarantees can be certified, and the criterion what their availability cost is elsewhere.

4 Structural Theory: When Composition Destroys Value

Two questions organize this section: when is coalition value guaranteed monotone, and, when it is not, whose presence destroys value, by how much, and through which credentials? Throughout, $v(S) = W(\hat{\sigma}_S)$ under full disclosure and $\Delta_x(S) \coloneqq v(S \cup \{x\}) - v(S)$. Full proofs are in the supplemental material; we give each argument's core below.

4.1 Monotonicity and Destructive Intermediaries

Throughout this subsection $\mathcal{G}$ is a rooted tree, so that Definition 3.11 is the recursion (4); Section 4.3 lifts what lifts. Recall (Definition 3.7) that an edge is preserving if it passes the whole of the child's set and filtering otherwise. The first theorem settles when coalition value can be trusted to be monotone at all—and shows that the answer depends on the operators alone, never on the shape of the tree.

Theorem 4.1 (Monotonicity Dichotomy)**.** *Let $\mathcal{G}$ be a rooted tree with operators drawn from $\mathcal{P}$ (in particular, from the four fundamental operators). The following are equivalent:*

(i) every edge of $\mathcal{G}$ is preserving, i.e., $\pi(\otimes) = \mathcal{C}$ (among the fundamental operators: $\cup$, $\rtimes_R$ for any R, or $\ltimes_{\mathcal{C}}$);

(ii) for every credential assignment σ and all coalitions $S \subseteq T$, we have $\hat{\sigma}_S \subseteq \hat{\sigma}_T$—hence $v(S) \leq v(T)$ for every monotone W.

Moreover, if (i) fails on some edge, there exists a credential assignment with all intrinsic sets nonempty and coalitions $S \subset T$ with $v(T) < v(S)$ under $W = |\cdot|$.

Proof sketch: (i)⇒(ii): add one actor x. Each child of x in S was a component root and now receives $\otimes(\hat{\sigma}_T(x), \sigma) \supseteq \sigma \cap \pi = \sigma$ (lower pass bound, $\pi = \mathcal{C}$); below it first arguments only grow and every $\otimes \in \mathcal{P}$ is monotone in its first argument (Lemma 3.1), so $\hat{\sigma}_T(a) \supseteq \hat{\sigma}_S(a)$ for every $a \in S$. Converse: if (a, b) is filtering with $z \notin \pi$, take $\sigma(a) = \{w\}$, $\sigma(b) = \{z, w\}$, $S = \{b\}$, $T = \{a, b\}$; exactness gives $z \notin \hat{\sigma}_T(b)$, so $v(T) = 1 < 2 = v(S)$. The gadget is local, so no restriction on topology can prevent it. □

For the assignment actually deployed, monotonicity is decided credential by credential, by where each copy sits:

Definition 4.1 (Protected holders)**.** *Let $\mathcal{G}$ be a rooted tree with operators in $\mathcal{P}$, and let $c \in \sigma(h)$. The holder h is* c-protected *if h is the root, or $c \in \pi(\otimes_{(p,h)})$ for the parent p of h, or $c \in \sigma(p)$ and p is c-protected. Equivalently, the run of consecutive holders of c ascending from h reaches the root or an edge that passes c. A credential is* protected *if every holder of it is.*

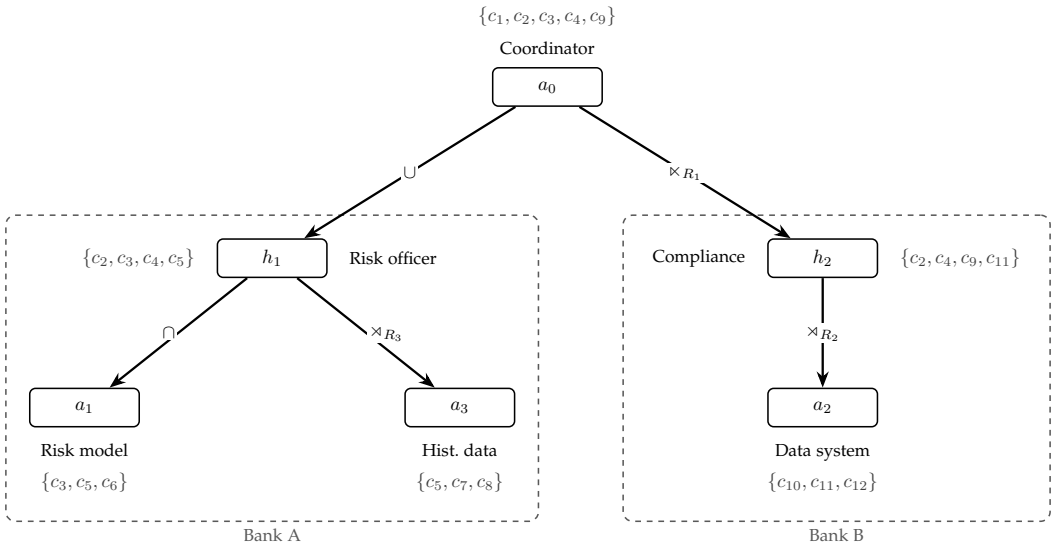


Fig. 1. The cross-border stress test: the actor tree with intrinsic credential sets and edge operators. The intersection edge $h_1 \to a_1$ is the only filter that destroys a uniquely held credential (c_6).

Proposition 4.1 (Credential protection)**.** *Let $\mathcal{G}$ be a rooted tree with operators in $\mathcal{P}$, and fix the assignment σ.*

(i) If $h \in S$ is a c-protected holder of c, then $c \in \hat{\sigma}_S(h)$; hence $c \in \hat{\sigma}_T$ for every coalition $T \ni h$.

(ii) $\hat{\sigma}_S \subseteq \hat{\sigma}_T$ for all $S \subseteq T$—no credential is ever lost, and v is monotone for every monotone W—if and only if every credential is protected.

(iii) Whether every credential is protected is decided in $O(|\mathcal{A}| \cdot |\mathcal{C}|)$ time by one pass from the root.

Proof sketch: (i) By exactness (Lemma 3.1), $c \in \hat{\sigma}_S(h)$ if and only if $c \in \hat{\sigma}_S(p) \cup \pi$ when the parent $p \in S$; induct up the run of holders. (ii) Every credential of $\hat{\sigma}_S$ has a holder in S, so (i) gives monotonicity. Conversely, the maximal ascending run $h_0, \ldots, h_k$ of holders above an unprotected h ends below a non-holder x with every edge filtering c: $\{h_0, \ldots, h_k\}$ holds c and $\{h_0, \ldots, h_k, x\}$ does not. (iii) One top-down pass. Theorem 4.1 is (ii) quantified over assignments: its gadget is an unprotected holder. □

When monotonicity fails, who can be the culprit? Not an actor that merely joins at the edge of the coalition:

Proposition 4.2 (Leaf joiners are harmless)**.** *Let $\mathcal{G}$ be a rooted tree. If x has no children in S, then $\Delta_x(S) = v(S \cup \{x\}) - v(S) \geq 0$, regardless of the operator on the edge connecting x to its parent—including intersection.*

Destruction thus requires a filtering edge *and* a child on the far side of it whose contribution the filter removes. The next theorem quantifies it exactly when no filtering edge lies below the filtered children within the coalition; without that condition the same accounting yields two-sided bounds (supplemental material), while the exact value of $\Delta_x(S)$ is always available in linear time (Proposition 6.1). The criterion's role is analytic: it says *which* credentials an intermediary destroys and *why*.

Theorem 4.2 (Destructive Intermediary Criterion)**.** *Let $\mathcal{G}$ be a rooted tree, $x \notin S$, $T = S \cup \{x\}$, and $W = |\cdot|$. Let $b_1, \ldots, b_m$ be the children of x in S whose edges are filtering, with pass filters $\pi_1, \ldots, \pi_m$, and suppose that each b_i is either isolated in F_S or connected downward in F_S only through union edges; the remaining children of x in S attach through preserving edges and may carry arbitrary subtrees. Let D be the set of members of S below some b_i, write*

$$\hat{\sigma}_T(x) = \begin{cases} \otimes_{(p,x)}(\hat{\sigma}_S(p), \sigma(x)), & \text{if } x\text{'s parent } p \in S, \\ \sigma(x), & \text{otherwise}, \end{cases}$$

and define the independent coverage *and the* loss set

$$N \coloneqq \bigcup_{a \in S \setminus (\{b_1, \ldots, b_m\} \cup D)} \hat{\sigma}_S(a) \;\cup\; \bigcup_{d \in D} \sigma(d),$$

$$L \coloneqq \Big(\bigcup_{i=1}^{m} \sigma(b_i)\Big) \setminus \Big(N \cup \hat{\sigma}_T(x) \cup \bigcup_{i=1}^{m} \big(\sigma(b_i) \cap \pi_i\big)\Big).$$

Then

$$\Delta_x(S) = \big|\hat{\sigma}_T(x) \setminus \hat{\sigma}_S\big| \;-\; |L|, \tag{6}$$

and L is exactly the set of credentials that x's presence destroys. In particular, x is destructive ($\Delta_x(S) < 0$) if and only if the credentials held only by its filtered children that neither survive their filters nor are resupplied by x outnumber x's genuinely novel contribution.

Proof sketch: Adding x changes $\hat{\sigma}$ in two places. Gain: x contributes $\hat{\sigma}_T(x)$, of which only the part outside $\hat{\sigma}_S$ is new; preserving subtrees only grow, and only by credentials of $\hat{\sigma}_T(x)$. Loss: a filtered child b_i, previously a component root contributing $\sigma(b_i)$, now receives $\otimes_{(x,b_i)}(\hat{\sigma}_T(x), \sigma(b_i))$, which by exactness retains $z \in \sigma(b_i)$ if and only if $z \in \hat{\sigma}_T(x) \cup \pi_i$; descendants on union edges keep their own $\sigma(d)$. So a credential disappears if and only if it was held only by filtered children and their heirs, survives no filter, and is not resupplied by x—the set L. □

4.2 A Worked Example: Cross-Border Regulatory Stress Test

Four worked examples accompany the theory; we keep in the main text the one that exercises all four operators and instantiates the structural results (Fig. 1).

Scenario: A regulator mandates a coordinated stress test across two banks. An autonomous coordinator agent a_0 assembles credentials from risk, compliance, and data actors at both institutions; no single actor can conduct the test alone. The credential universe $\mathcal{C} = \{c_1, \ldots, c_{12}\}$ comprises: c_1 stress test coordination authority, c_2 regulatory reporting certification, c_3 risk assessment methodology, c_4 AML clearance, c_5 Bank A internal systems access, c_6 VaR model certification, c_7 backtesting authorization, c_8 historical loss database access, c_9 cross-organizational data sharing agreement, c_{10} Bank B customer exposure data, c_{11} sanctions screening certification, and c_{12} Bank B internal audit access.

Actors, credentials, and edges: The actor graph is a tree rooted at a_0, with Bank A's risk function on the left and Bank B's compliance function on the right; Fig. 1 gives every actor's role, intrinsic credential set, and edge operator, including the relevance filters $R_1 = \{c_2, c_4, c_{11}\}$, $R_2 = \{c_9\}$, and $R_3 = \{c_3, c_5\}$.

Headline computation: Under shared-outcome evaluation with $W = |\cdot|$, the grand coalition achieves $v(N) = 11$: every credential except c_6, the risk model's unique VaR certification, which the intersection edge $h_1 \to a_1$ filters away. Removing the risk officer disconnects a_1 and a_3, reverting their contributions to their intrinsic sets, and $v(N \setminus \{h_1\}) = 12 > v(N)$: the grand coalition is not optimal, and h_1 has negative marginal contribution in 12 of the 16 coalitions in which its arrival re-attaches a filtered child. Theorem 4.2 derives these numbers by substitution: for $S = N \setminus \{h_1\}$ the gain is empty ($\hat{\sigma}_T(h_1) = \sigma(a_0) \cup \sigma(h_1)$ is already covered) and the loss set is $L = \{c_6\}$, the credential only the filtered child a_1 holds; for $S = \{a_1\}$ the gain is $\{c_2, c_4\}$ against the same loss, $+1$. At-target evaluation at the risk model tells the same story: $\Delta_{h_1} = |\{c_3, c_5\}| - |\sigma(a_1)| = -1$ under intersection, and $+4$ once the edge $h_1 \to a_1$ carries a left join $\ltimes_R$ with $c_6 \in R$—the oversight edge is retained, the unique credential survives.

The full development, the coalition game with Shapley analysis, the operator-redesign remedy, and the exhaustive computation of the values of all 63 nonempty coalitions of this example are in the supplemental material.

4.3 Beyond Trees: Lifting the Structural Results

Definition 3.11 already defines effective sets on arbitrary directed graphs; the structural theorems of Section 4.1 were stated for trees. The strategic results of Section 5 are graph-free, and the hardness results of Section 6 transfer *a fortiori* from depth-one trees. We now lift the dichotomy, the protection criterion, and leaf-harmlessness, and identify exactly when the choice of origins matters. Destructiveness itself is general: $\Delta_x(S)$ is computable on any directed graph (Proposition 3.2), and the next theorem says when it can occur; tree-bound is only its closed-form accounting. The condition "every edge preserving" acquires one refinement: an edge into an origin cannot filter anything the origin already holds.

Theorem 4.3 (Monotonicity Dichotomy on Directed Graphs)**.** *Let $\mathcal{G}$ be any directed graph with operators drawn from $\mathcal{P}$ and origin set O, under the semantics of Definition 3.11. The following are equivalent:*

(i) every edge (p, a) whose head a is not an origin ($a \notin O$) is preserving;

(ii) for every credential assignment σ and all coalitions $S \subseteq T$: $\hat{\sigma}_S \subseteq \hat{\sigma}_T$.

Moreover, if (i) fails on some edge, there exist a credential assignment with all intrinsic sets nonempty and coalitions $S \subset T$ with $v(T) < v(S)$ under $W = |\cdot|$.

Proof sketch: (i)⇒(ii): under (i), $\Theta_U(\bot)(a) \supseteq \sigma(a)$ for every a (origins by the base term, all others through a preserving parent edge), and the upper bound of Lemma 3.1 gives $\hat{\sigma}_U \subseteq \bigcup_U \sigma$; so $\hat{\sigma}_U = \bigcup_{a \in U} \sigma(a)$, monotone in U. Converse: the two-actor gadget of Theorem 4.1, with an induction showing that z enters no iterate. □

Proposition 4.1 lifts to acyclic graphs verbatim—"every parent" in place of "the parent," origins as the base case—and stays sufficient on cyclic graphs, where it stops being exact (a three-cycle of holders filtering c from one another, plus one union edge, protects no holder yet never loses c; supplemental material). There the exact question is hard, for a reason unrelated to Theorem 6.1: not netting gains against losses, but closing a consistent loop of filters around every holder of one credential.

Proposition 4.3 (Monotonicity of a deployment is coNP–complete on cyclic graphs)**.** MONOTONE-DEPLOYMENT*—given $\mathcal{G}$, its operators, origins O, and σ, is $\hat{\sigma}_S \subseteq \hat{\sigma}_T$ for all $S \subseteq T$?—is decidable in $O((|\mathcal{A}| + |E|) \cdot |\mathcal{C}|)$ time on acyclic graphs (supplemental material) and coNP-complete on directed graphs with cycles, already with $|\mathcal{C}| = 1$, $O = \emptyset$, operators $\cap$ and $\cup$ only, and every actor holding the credential. In the same setting* DESTRUCTIVE-ACTOR *is NP-complete; hence the fixed-parameter tractability in $|\mathcal{C}|$ of Proposition 6.3 does not extend to cyclic graphs unless* P = NP.

Proof sketch: A coalition loses c if and only if every holder in it has a parent in it and none a passing one. Reduce SAT: a $\cap$-necklace of clause and literal actors with $\cup$-edges between complementary literals; losing c means going around the necklace choosing one literal per clause and no complementary pair—a satisfying assignment. □

Aggregation is not a boundary: any monotone aggregator between intersection and union in place of the union over parents in (3) leaves the preserving regime, hence the dichotomy, unchanged under shared-outcome evaluation (supplemental material; conjunctive aggregation can destroy value only *at a target*). Leaf-harmlessness lifts with "leaf" read as "sink": an actor with no edge into the coalition cannot filter anyone.

Proposition 4.4 (Sinks Are Harmless)**.** *Let $x \notin S$ have no edge into S, so that x is a sink of $F_{S \cup \{x\}}$. Then $\hat{\sigma}_{S \cup \{x\}} \supseteq \hat{\sigma}_S$ for every credential assignment, origin set, and operator assignment; equivalently, $\Delta_x(S) \geq 0$ under shared-outcome evaluation with monotone W.*

On a general graph the query origin must be designated; when the designation matters is again a pass-filter condition, now per actor.

Proposition 4.5 (When the Origin Matters)**.** *Fix $\mathcal{G}$ and its operators, and consider the grand coalition $S = \mathcal{A}$. The least solution is the same for every origin set $O \supseteq \mathrm{src}(\mathcal{G})$ and every*

TABLE 2
Payoffs (u_a, u_b) in the proof of Proposition 5.1: every profile admits a profitable deviation.

	$\delta_{b\to a}=\emptyset$	$\delta_{b\to a}=\{y\}$
$\delta_{a\to b}=\emptyset$	$(0,\ 0)$	$(0,\ 9)$
$\delta_{a\to b}=\{x\}$	$(-1,\ 10)$	$(9,\ 9)$

credential assignment σ if and only if every non-source actor a satisfies the incoming pass cover *condition*

$$\bigcup_{p\in\mathrm{par}(a)} \pi\big(\otimes_{(p,a)}\big) = \mathcal{C}.$$

Proposition 4.5 turns the tree-rooting assumption into a theorem: the origin is the anchor the filtering regime requires, and on a rooted tree it is structurally forced. In sum: semantics, evaluation, the dichotomy, sink-harmlessness, and all hardness and sampling results hold on arbitrary directed graphs; the protection audit holds up to acyclic graphs; the exact criterion (two-sided bounds beyond trees, where a lost credential can be resupplied along another walk), the fixed-parameter algorithm, and target-directed synthesis remain tree-bound.

5 Strategic Layer: Disclosure and Coalition Value

The algebra determines what *could* propagate, not whether actors disclose it—revealing a risk model exposes competitive intelligence. We model disclosure as a game on an edge and coalition value as a cooperative game generated by the algebra.

5.1 The Disclosure Game and Its Equilibria

Definition 5.1 (Disclosed Credential Set)**.** *For an actor $a \in \mathcal{A}$, the* disclosed credential set *toward actor b is a subset $\delta_{a\to b} \subseteq \sigma(a)$ of a's intrinsic credential set (Definition 3.3) chosen strategically by a. The operator $\otimes_{a,b}$ then acts on disclosed sets rather than intrinsic sets:*

$$\hat{\sigma}_{a,b} = \otimes_{a,b}(\delta_{a\to b},\ \delta_{b\to a}) \tag{7}$$

Definition 5.2 (Disclosure Game)**.** *The* disclosure game *on edge $(a,b) \in E$ is a simultaneous game $\Gamma_{a,b} = (N, \{S_i\}, \{u_i\})$ where:*

(i) $N = \{a, b\}$; the strategy sets are $S_a = \mathcal{P}(\sigma(a))$ and $S_b = \mathcal{P}(\sigma(b))$, the subsets of intrinsic credentials each player may disclose.

(ii) The utility of player $i \in \{a, b\}$ is

$$u_i(\delta_{a\to b}, \delta_{b\to a}) = V_i(\otimes_{a,b}(\delta_{a\to b}, \delta_{b\to a})) - D_i(\delta_{i\to j}) \tag{8}$$

where $V_i : \mathcal{P}(\mathcal{C}) \to \mathbb{R}_{\geq 0}$ is the value function *measuring the benefit to i from the effective credential set, and $D_i : \mathcal{P}(\sigma(i)) \to \mathbb{R}_{\geq 0}$ is the* disclosure cost *measuring the cost to i of revealing credentials.*

Proposition 5.1 (Non-existence of pure-strategy equilibria)**.** *There exist disclosure games $\Gamma_{a,b}$—with the union operator, monotone value functions, and additive disclosure costs—that admit no pure-strategy Nash equilibrium.*

Proof: Let $\mathcal{C} = \{x, y\}$, $\sigma(a) = \{x\}$, $\sigma(b) = \{y\}$, union on the edge, unit disclosure costs, $V_a(T) = 10$ if $T = \{x, y\}$ and 0 otherwise (complements for a), $V_b(T) = 10$ if $T \neq \emptyset$ and 0 otherwise (substitutes for b). The payoffs (u_a, u_b) are given in Table 2: each profile admits a profitable deviation (b up, a up, b down, a down), a matching-pennies best-response cycle. Mixed equilibria exist by Nash's theorem; sequential disclosure restores pure existence by backward induction. □

5.2 Equilibrium Classification

Definition 5.3 (Modularity classes)**.** *A set function $F : \mathcal{P}(\mathcal{C}) \to \mathbb{R}$ is*

(i) modular *if $F(T) = F(\emptyset) + \sum_{c\in T} w(c)$ for some weights $w : \mathcal{C} \to \mathbb{R}$;*

(ii) supermodular *if $F(X\cup Y)+F(X\cap Y) \geq F(X)+F(Y)$ for all $X, Y \subseteq \mathcal{C}$; equivalently, if the marginal value $F(T \cup \{c\}) - F(T)$ is nondecreasing in T for every $c \notin T$;*

(iii) submodular *if the reverse inequalities hold.*

Supermodular value means the credentials are complements*, submodular that they are* substitutes*.*

Proposition 5.2 (No-disclosure deadlock under intersection)**.** *Let the edge operator be intersection, and let the disclosure costs satisfy $D_i(\emptyset) = 0$ and $D_i \geq 0$. Then for* all *value functions V_a, V_b, the profile $(\emptyset, \emptyset)$ is a pure-strategy Nash equilibrium of $\Gamma_{a,b}$. If moreover $D_i \equiv 0$ and each V_i is monotone with respect to set inclusion, full disclosure $(\sigma(a), \sigma(b))$ is also a pure-strategy equilibrium.*

Under intersection, mutual non-disclosure is thus an equilibrium but a Pareto-dominated one: the disclosure game on an intersection edge is a weakest-link coordination game [18].

Proposition 5.3 (Modular preferences: equilibria under all four operators)**.** *Let the edge operator be any of the four fundamental operators, and let all V_i and D_i be modular with $V_i(\emptyset) = D_i(\emptyset) = 0$ (hence with nonnegative weights $w_i(c) = V_i(\{c\})$ and $d_i(c) = D_i(\{c\})$). Then $\Gamma_{a,b}$ admits a pure-strategy Nash equilibrium.*

Theorem 5.1 (Strategic complementarity under union)**.** *Let the edge operator be union, let the intrinsic credential sets be disjoint ($\sigma(a) \cap \sigma(b) = \emptyset$), let each V_i be supermodular, and let each D_i be submodular (in particular, modular costs qualify). Then $\Gamma_{a,b}$ is a supermodular game on the lattices $(\mathcal{P}(\sigma(a)), \subseteq)$ and $(\mathcal{P}(\sigma(b)), \subseteq)$; consequently it admits a pure-strategy Nash equilibrium, and the equilibrium set contains a largest and a smallest equilibrium [19].*

The existence argument is Tarski's fixed-point theorem [17] applied to the best-response map on the lattice of disclosure profiles—the same tool that underlies Proposition 3.2, where monotonicity is unconditional; here it is exactly strategic complementarity, and Proposition 5.1 is what a non-monotone best response looks like. The result transfers to join edges by reduction to union on the ground set $\sigma(b) \cap R$.

Proposition 5.4 (Common interests: the disclosure game is a potential game)**.** *Suppose $V_a = V_b = V$ for an arbitrary value function V, with arbitrary disclosure costs and an* arbitrary

TABLE 3
Existence of pure-strategy disclosure equilibria, by the operator on the edge and the preference structure of the two players.

Preference structure	Intersection $\cap$	Union $\cup$	Joins $\ltimes_R$, $\rtimes_R$
Arbitrary V_i	exists (Prop. 5.2)	may fail (Prop. 5.1)	may fail
Modular V_i, D_i	exists	exists	exists (Prop. 5.3)
Supermodular V_i (disjoint holdings)	exists	exists (Thm. 5.1)	exists
Common interest ($V_a = V_b$)	exists	exists	exists (Prop. 5.4)

edge operator $\otimes$. Then $\Gamma_{a,b}$ is an exact potential game [20] with potential

$$P(\delta_{a\to b}, \delta_{b\to a}) = V(\otimes(\delta_{a\to b}, \delta_{b\to a})) - D_a(\delta_{a\to b}) - D_b(\delta_{b\to a}),$$

and any maximizer of P is a pure-strategy Nash equilibrium; in particular, pure equilibria exist.

Table 3 summarizes the classification. Proposition 5.1 is tight: non-existence requires an expansive operator and strictly opposite complementarity on the two sides, and it transfers to joins by $R = \mathcal{C}$. Both-submodular preferences with multi-credential holdings remain open.

5.3 Coalition Value and Attribution

Definition 5.4 (Credential Coalition Game)**.** *Given an actor graph $\mathcal{G} = (\mathcal{A}, E)$, the* credential coalition game *is a pair $(\mathcal{A}, v)$ where the characteristic function $v : \mathcal{P}(\mathcal{A}) \to \mathbb{R}_{\geq 0}$ assigns to each coalition $S \subseteq \mathcal{A}$ the value of the effective credential set achievable by the actors in S collaborating optimally:*

$$v(S) = \max_{\substack{\delta_{i\to j} \subseteq \sigma(i) \\ \forall (i,j) \in E_S}} W(\hat{\sigma}_S) - \sum_{i \in S} D_i(\delta_i) \tag{9}$$

where $E_S = E \cap (S \times S)$ is the set of edges internal to S, $\hat{\sigma}_S$ is the effective credential set computed over the subgraph induced by S with optimal disclosure, W is the valuation of Definition 3.12, and $\delta_i \coloneqq \bigcup_j \delta_{i\to j} \subseteq \sigma(i)$ is the set *of credentials actor i discloses on any of its edges in S. A credential revealed once is revealed: the cost $D_i(\delta_i)$ is charged per credential, not per edge, so disclosing the same credential on a second edge is free; the bilateral V_i are subsumed by W at the coalition level.*

Definition 5.5 (Marginal Contribution)**.** *The* marginal contribution *of actor a to coalition S (where $a \notin S$) is:*

$$\Delta_a(S) = v(S \cup \{a\}) - v(S) \tag{10}$$

Definition 5.6 (Shapley Value [21])**.** *The* Shapley value *of actor $a \in \mathcal{A}$ is:*

$$\phi_a = \sum_{S \subseteq \mathcal{A} \setminus \{a\}} \frac{|S|!\,(|\mathcal{A}| - |S| - 1)!}{|\mathcal{A}|!}\, \Delta_a(S) \tag{11}$$

The equilibrium results above are edge-level (their proofs use locality, which fails across a path); the maximizing profile in (9) exists by finiteness and is a cooperative optimum, not an equilibrium (supplemental material). Under union edges with a superadditive W the game is superadditive and the grand coalition efficient (supplemental material); under cardinality it is monotone but merely subadditive on overlapping holdings; under filtering edges monotonicity itself fails, as in the stress-test example, where $v(N) < v(N \setminus \{h_1\})$. The Shapley value attributes *authorization potential*: negative values arise structurally, from filtering edges, and are meaningful. A Vickrey–Clarke–Groves mechanism elicits truthful disclosure and serves as an audit benchmark rather than a deployable protocol (budget imbalance, collusion, Sybil splitting, and tension with disclosure minimality; supplemental material).

6 Computational Complexity and Synthesis

Proposition 6.1 (Coalition evaluation is linear on trees)**.** *Given a rooted tree $\mathcal{G}$, an operator assignment, a credential assignment, and a coalition S, the coalition effective set $\hat{\sigma}_S$ and the value $v(S) = |\hat{\sigma}_S|$ can be computed in time $O(|S| \cdot |\mathcal{C}|)$. Consequently, any single marginal contribution $\Delta_x(S)$ is computable in time $O(n \cdot |\mathcal{C}|)$.*

On arbitrary directed graphs evaluation is polynomial rather than linear (Proposition 3.2); every membership and sampling result below therefore holds on any graph, and every hardness result holds *a fortiori*.

Definition 6.1 (The DESTRUCTIVE-ACTOR problem)**.** Input: *a rooted tree $\mathcal{G}$, an operator assignment drawn from the four fundamental operators, a credential assignment σ, and a designated actor x.* Question: *does there exist a coalition $S \subseteq \mathcal{A} \setminus \{x\}$ with $\Delta_x(S) < 0$?*

Theorem 6.1 (Detecting destructiveness is NP-complete)**.** *The* DESTRUCTIVE-ACTOR *problem is NP-complete. Hardness holds already for trees of depth one using only the operators $\cap$ and $\cup$, and therefore transfers to every graph class containing such trees; combined with Proposition 3.2, the problem is NP-complete on arbitrary directed graphs as well. Consequently, deciding that a given actor is* never *destructive is coNP-complete.*

Proof sketch: Membership: a coalition S is a certificate (Proposition 6.1). Hardness, from RED-BLUE SET COVER [22] (blue set B, red set R, sets Q_j, budget $k \leq |R|-1$): root x holds $M = |R|+1$ copies of each blue element plus $g_0 = |R| - k - 1$ dummies; an intersection child c holds R; union children q_j hold the blue copies of $Q_j \cap B$ and $Q_j \cap R$. All children are siblings, so Theorem 4.2 gives, for $S = \{c\} \cup \{q_j : j \in J\}$, $\Delta_x(S) = M \cdot \mathrm{ub}(J) + g_0 - \mathrm{ur}(J)$ (ub, ur: uncovered blue and red counts). If any blue is uncovered, $\Delta \geq M > \mathrm{ur}$; otherwise $\Delta < 0$ if and only if at most k reds are covered. Coalitions without c are never destructive. □

Proposition 6.2 (Counting destructive coalitions is #P-complete)**.** *Given an instance as in Definition 6.1, computing $|\{S \subseteq \mathcal{A} \setminus \{x\} : \Delta_x(S) < 0\}|$ is #P-complete. As with Theorem 6.1, hardness on depth-one trees transfers to every containing graph class.*

Proposition 6.3 (The hardness is in the actors, not the credentials)**.** *On rooted trees,* DESTRUCTIVE-ACTOR *is decidable in time $2^{O(|\mathcal{C}|)} \cdot |\mathcal{A}|$; the number of coalitions S with $\Delta_x(S) < 0$ and the exact Shapley value ϕ_x are computable in time $2^{O(|\mathcal{C}|)} \cdot \mathrm{poly}(|\mathcal{A}|)$. Thus all three problems are fixed-parameter tractable in the number of credentials: the hardness of Theorem 6.1 and Proposition 6.2 lies in the number of actors.*

Proof sketch: On a tree an actor's effective set depends only on its *context* (the parent's effective set, or none if the parent is absent) and its own set, so given the context the children's subtrees are independent, and only actors below x see different contexts with and without x. A bottom-up dynamic program over pairs of contexts (X^-, X^+) computes, per subtree, the achievable pairs (Y^-, Y^+) of unions of effective sets under S and $S \cup \{x\}$; there are $2^{O(|\mathcal{C}|)}$ of each, and x is destructive if and only if the root admits $|Y^+| < |Y^-|$. Multiplicity tables give the count; tracking $|S|$ gives ϕ_x. □

Proposition 6.4 (Shapley values: exact hardness, efficient approximation)**.** *Exact Shapley computation is #P-hard for succinct cooperative games in general [23]; here the value oracle is polynomial (Proposition 6.1), so for any $\varepsilon > 0$ and $\delta \in (0, 1)$, averaging the marginal contributions of x over $m \geq 2|\mathcal{C}|^2 \ln(2/\delta)/\varepsilon^2$ uniformly random permutations yields $\hat{\phi}_x$ with $\Pr[|\hat{\phi}_x - \phi_x| \geq \varepsilon] \leq \delta$, in total time $O(m \cdot n \cdot |\mathcal{C}|)$ [24].*

Coalition synthesis: Audit asks whether a given coalition is authorized; synthesis asks for the cheapest one.

Definition 6.2 (Minimum-Cost Authorizing Coalition (MCAC))**.** Input: *a rooted tree $\mathcal{G}$ with operator assignment, credential assignment σ, activation costs $c : \mathcal{A} \to \mathbb{Q}_{\geq 0}$, and a requirement $Q \subseteq \mathcal{C}$.* Shared-outcome variant (MCAC): *find $S \subseteq \mathcal{A}$ minimizing $c(S) = \sum_{a \in S} c(a)$ subject to $\hat{\sigma}_S \supseteq Q$.* Target variant (MCAC-T): *additionally given a target actor t, find $S \ni t$ minimizing $c(S)$ subject to $\hat{\sigma}_S(t) \supseteq Q$.*

Proposition 6.5 (Preserving regime: exactly weighted set cover)**.** *If every edge of $\mathcal{G}$ is preserving, then $\hat{\sigma}_S = \bigcup_{a \in S} \sigma(a)$ for every coalition S, and* MCAC *is equivalent to weighted* SET COVER *with universe Q, sets $\sigma(a) \cap Q$, and weights $c(a)$. Consequently* MCAC *is NP-hard already in the preserving regime [25], the greedy algorithm achieves an $H(|Q|)$ approximation [26], and no polynomial-time algorithm can do asymptotically better unless* $\mathrm{P} = \mathrm{NP}$ *[27].*

Proposition 6.6 (Target variant: polynomial on trees for all four operators)**.** MCAC-T *is solvable in time $O(d \cdot |\mathcal{C}|)$, where d is the depth of the target t—for arbitrary mixtures of the four fundamental operators.*

Proposition 6.7 (Filtering regime: the coverage structure collapses)**.** *Let $f(S) := |Q \cap \hat{\sigma}_S|$ denote the coverage function of an* MCAC *instance. In the presence of a filtering edge:*

(i) f need not be monotone—a feasible coalition can have infeasible supersets;

(ii) f need not be submodular: marginal coverage can strictly increase along a chain of coalitions;

(iii) the decision version of MCAC remains NP-complete (it contains the instances of Proposition 6.5), but the monotone-submodular structure behind the greedy guarantee is absent, and the approximability of MCAC with filtering edges is open.

The complexity landscape: Auditing a deployed configuration is cheap: values, marginal contributions, and uniquely covered credentials are linear-time. Certifying safety is hard—NP-complete to detect a possibly destructive actor, coNP-complete to certify one safe, already at depth one with two operators—so global guarantees come only by operator regime (Theorem 4.1), never by topology. Two weaker certificates stay cheap: that a deployment can never *lose* a credential is linear up to acyclic graphs (Proposition 4.1), and with $|\mathcal{C}|$ fixed, detection, counting, and exact attribution are polynomial on trees (Proposition 6.3); both cheap routes end at cycles (Proposition 4.3), and otherwise attribution is statistical. Synthesis follows the dichotomy's fault line: greedily approximable in the preserving regime, polynomial at a target on trees, without submodular structure once a filter appears.

Computational evaluation: On random trees (supplemental material, Section S11; code in the repository cited below), one audit $\Delta_x(S)$ takes 0.5 ms at 10^3 actors and 0.18 s at 10^5, independently of $|\mathcal{C}|$; exact detection (Proposition 6.3) takes 8 ms at $|\mathcal{C}| = 8$ and 0.3 s at $|\mathcal{C}| = 12$ for $n = 200$. At filter density 0.6 about 7% of actors are destructive in some coalition, and redundancy helps exactly as Proposition 4.1 predicts: protected copies cut destructive pairs tenfold; random copies triple the ever-destructive actors.

7 DISCUSSION

7.1 Design Principles

Four operational principles follow; each has a well-studied neighbor, named first, and a channel the algebra adds, named second.

Oversight placement has an algebraic price, and leaves are free. The behavioral costs of human oversight—vigilance decay, automation bias—are well documented [28]. The algebra reveals a second, independent channel that operates with perfectly attentive humans: the cost is carried by the overseer's position and edge semantics. A leaf overseer is algebraically free under shared-outcome evaluation, whatever its operator (Proposition 4.2); any intermediary placement is priced by (6). The rule is not "less oversight" but oversight at the leaves where possible and priced elsewhere; when an overseer must sit mid-tree, replacing an intersection edge by a join that keeps the oversight role while filtering less is the remedy, and it rests on a theorem. Nothing here is specific to humans: an AI intermediary on the same edge is priced identically.

Redundancy protects only from protected positions. In reliability theory the Barlow–Proschan importance measure is a Shapley value of a coherent—hence monotone—reliability game [29]. In our non-monotone setting the loss set L of Theorem 4.2 is the criticality notion with a different threat: redundancy protects a credential not against component *failure* but against the *presence* of a filtering intermediary. What protects is not the number of holders but a protected one (Definition 4.1): such a copy guards the credential in every coalition containing it, whereas a copy behind an edge that filters it guards nothing and is a new target, since it creates

TABLE 4
Question classes made expressible, what blocked each, and their status here.

Question class	Why previously inexpressible	Status here
Harm. Can adding actor x reduce what a configuration is authorized to do?	Pre-answered by the monotonic-restriction axiom.	Decidable per configuration in linear time (Prop. 6.1); characterized by Thm. 4.2.
Placement. Where should a human overseer sit?	The human is the axiomatic root; a root has no alternative positions.	Leaves free (Prop. 4.2); any other placement priced by (6).
Attribution. What does an actor's presence cost, and who bears it?	Policy expressions have no actors; consent structures cannot generate negative positional value.	Shapley attribution over the algebra-generated game; negative values meaningful, sampled (Prop. 6.4).
Diagnosis. Which credential does an oversight arrangement destroy?	Scalar trust has no content; permission bits carry no sets to lose.	Read off the loss term of (6) (c_6 in the stress-test example).
Doctrine. Is mutual non-disclosure rational, and efficient?	Disclosure was not a strategic variable over credential sets under operator semantics.	An equilibrium, yes; efficient, often not (Prop. 5.2); remedied by sequencing.
Certification. Can a regulator guarantee safety by mandating structure?	Operators were not a design variable; value-destruction safety is empty in monotone worlds.	By operator regime only, never by topology (Thm. 4.1); no-loss per deployment linear up to acyclic graphs (Prop. 4.1); per actor coNP-complete (Thm. 6.1), FPT in $\lvert\mathcal{C}\rvert$ on trees (Prop. 6.3).

a coalition in which the credential is lost (Proposition 4.1). Protection is a linear-time audit that certifies what no per-actor test can cheaply: a protected deployment never loses a credential, whereas certifying that one actor never nets a loss is coNP-complete (Theorem 6.1). The hardness is in the netting, not in finding the exposure.

Audit is cheap, certification is hard—the safety-analysis tradition, extended. Safety analysis of protection systems begins with Harrison, Ruzzo, and Ullman (HRU) [30], who showed the safety problem undecidable in their access-matrix model; its complexity has since been mapped for role-based models [31]. Theorem 6.1 continues this tradition into a question monotone models cannot pose: whether an actor's presence can destroy value. The stance mirrors access-control experience: audit exactly, certify by operator regime only, attribute by sampling.

Mutual non-disclosure as coordination failure. Rational lock-in at the inferior equilibrium of a weakest-link game is classical [18], and we claim no novelty for it. The algebra contributes the identification: on intersection edges the never-disclose-without-necessity maxim is not prudence but the Pareto-dominated equilibrium of Proposition 5.2, dominated in the players' own payoffs with disclosure costs charged. This concerns task availability, not confinement: zero-trust architecture in the NIST sense [32] verifies every request and so presupposes disclosure; the deadlock is the profile in which nothing is disclosed. The known cure, sequencing, is sequential disclosure, whose backward induction restores efficient pure play: disclosure order is a governance lever.

None of the four principles claims novelty for the phenomenon it names; each is documented in a neighboring discipline, cited above. What is new is that the same phenomena have a second, independent cause inside authorization semantics, and that this cause admits an exact statement: a formula (Theorem 4.2), a characterization (Theorem 4.1, Proposition 4.1), or a complexity boundary (Theorem 6.1). The two causes add rather than substitute: a perfectly attentive human behind the intersection edge still destroys c_6 in the stress-test example. Each principle thus comes with a checkable action: place overseers at leaves, or price their placement by (6); duplicate uniquely covered credentials into protected positions; audit every configuration exactly, and demand operator-regime rather than topological guarantees; and fix the disclosure order before the incentives.

7.2 Questions the Framework Makes Expressible

Table 4 lists six *classes of questions* that become well-posed once actors owning content, operators that can destroy it, and coalitional counterfactuals are present simultaneously; each prior tradition supplies one or two of the three.

7.3 Open Problems

Exact destructiveness accounting beyond trees; fixed-parameter tractability of DESTRUCTIVE-ACTOR on acyclic graphs; target-directed synthesis beyond trees; approximability of MCAC with filtering edges; both-submodular disclosure equilibria; endogenous choice of relevance filters R as a two-stage game; stable attribution over a drifting topology; and the dynamics of repeated re-evaluation on cyclic graphs, where iteration from the intrinsic state converges whenever every actor's incoming pass filters cover $\mathcal{C}$ but can otherwise oscillate—an authorization *flapping*, a relative of the route oscillations of the stable paths problem [33], whose convergence criterion is open.

8 CONCLUSION

We have presented a credential algebra in which human and AI actors are structural peers, with effective authorization computed by set operators on a directed graph and coalition value generated by the algebra itself. Monotonicity of coalition value is characterized exactly, at the level of operators (every edge passes its subordinate endpoint unfiltered, on

trees and, under the least-solution semantics, on arbitrary directed graphs) and at the level of a deployed assignment (every holder protected); where it fails, the destructive-intermediary criterion prices each actor's presence in closed form, with leaves and sinks provably harmless. Auditing is linear, certification coNP-complete yet fixed-parameter tractable in the number of credentials on trees, and hard again on cycles; attribution is approximable by sampling. Every question answered here is unposable under monotonic restriction; the contribution is to make them askable and to answer them, up to proofs of which are computationally answerable at all.

Two directions stand out beyond Section 7.3: embedding the algebra into a Datalog-based credential language, whose grounded semantics it already shares, so that destructiveness audits reach deployed credential chains; and instrumenting the criterion, since how often deployed agent workflows place a filtering intermediary on an unprotected credential is now a precise empirical question.

ACKNOWLEDGMENTS

The author used Anthropic's Claude throughout the preparation of this manuscript for drafting and editing text and LaTeX, checking calculations, and internet search; all mathematical statements and proofs were independently verified by the author, who takes full responsibility for the content. This research received no specific grant from any funding agency. Code reproducing every number in this paper: https://github.com/yryabov/game-of-peers, archived at https://doi.org/10.5281/zenodo.23002446.

REFERENCES

[1] T. South, S. Marro, T. Hardjono, R. Mahari, C. D. Whitney, D. Greenwood, A. Chan, and A. Pentland, "Authenticated delegation and authorized AI agents," *arXiv preprint arXiv:2501.09674*, 2025.

[2] C. Ellison, B. Frantz, B. Lampson, R. Rivest, B. Thomas, and T. Ylonen, "SPKI certificate theory," RFC 2693, Internet Engineering Task Force, September 1999.

[3] "Authenticated workflows: A systems approach to protecting agentic AI," *arXiv preprint arXiv:2602.10465*, 2026.

[4] I. Terrucha, E. Fernández Domingos, R. Suchon, F. C. Santos, P. Simoens, and T. Lenaerts, "Humans program artificial delegates to accurately solve collective-risk dilemmas but lack precision," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 122, no. 25, e2319942121, 2025.

[5] C. M. de Melo, S. Marsella, and J. Gratch, "Human cooperation when acting through autonomous machines," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 8, pp. 3482–3487, 2019.

[6] R. P. Gilles, G. Owen, and R. van den Brink, "Games with permission structures: The conjunctive approach," *International Journal of Game Theory*, vol. 20, no. 3, pp. 277–293, 1992.

[7] P. Bonatti, S. De Capitani di Vimercati, and P. Samarati, "An algebra for composing access control policies," *ACM Transactions on Information and System Security*, vol. 5, no. 1, pp. 1–35, 2002.

[8] D. Wijesekera and S. Jajodia, "A propositional policy algebra for access control," *ACM Transactions on Information and System Security*, vol. 6, no. 2, pp. 286–325, 2003.

[9] N. Li, J. C. Mitchell, and W. H. Winsborough, "Design of a role-based trust-management framework," in *Proc. 2002 IEEE Symposium on Security and Privacy*, pp. 114–130, 2002.

[10] M. Blaze, J. Feigenbaum, and J. Lacy, "Decentralized trust management," in *Proc. 1996 IEEE Symposium on Security and Privacy*, pp. 164–173, 1996.

[11] N. Li and J. C. Mitchell, "Understanding SPKI/SDSI using first-order logic," *International Journal of Information Security*, vol. 2, no. 1, pp. 47–64, 2003.

[12] OASIS, "eXtensible Access Control Markup Language (XACML) Version 3.0," OASIS Standard, January 2013.

[13] R. B. Myerson, "Graphs and cooperation in games," *Mathematics of Operations Research*, vol. 2, no. 3, pp. 225–229, 1977.

[14] H. Huang et al., "A novel zero-trust identity framework for agentic AI: Decentralized authentication and fine-grained access control," *arXiv preprint arXiv:2504.13538*, 2025.

[15] I. Pinyol and J. Sabater-Mir, "Computational trust and reputation models for open multi-agent systems: A review," *Artificial Intelligence Review*, vol. 40, no. 1, pp. 1–25, 2013.

[16] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *IEEE Computer*, vol. 29, no. 2, pp. 38–47, 1996.

[17] A. Tarski, "A lattice-theoretical fixpoint theorem and its applications," *Pacific Journal of Mathematics*, vol. 5, no. 2, pp. 285–309, 1955.

[18] J. B. Van Huyck, R. C. Battalio, and R. O. Beil, "Tacit coordination games, strategic uncertainty, and coordination failure," *American Economic Review*, vol. 80, no. 1, pp. 234–248, 1990.

[19] D. M. Topkis, *Supermodularity and Complementarity*. Princeton, NJ: Princeton University Press, 1998.

[20] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[21] L. S. Shapley, "A value for n-person games," in *Contributions to the Theory of Games*, vol. II (H. W. Kuhn and A. W. Tucker, eds.), Annals of Mathematics Studies, no. 28, Princeton University Press, pp. 307–317, 1953.

[22] R. D. Carr, S. Doddi, G. Konjevod, and M. V. Marathe, "On the red-blue set cover problem," in *Proc. 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 345–353, 2000.

[23] X. Deng and C. H. Papadimitriou, "On the complexity of cooperative solution concepts," *Mathematics of Operations Research*, vol. 19, no. 2, pp. 257–266, 1994.

[24] J. Castro, D. Gómez, and J. Tejada, "Polynomial calculation of the Shapley value based on sampling," *Computers & Operations Research*, vol. 36, no. 5, pp. 1726–1730, 2009.

[25] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations* (R. E. Miller and J. W. Thatcher, eds.), pp. 85–103, Plenum Press, 1972.

[26] V. Chvátal, "A greedy heuristic for the set-covering problem," *Mathematics of Operations Research*, vol. 4, no. 3, pp. 233–235, 1979.

[27] U. Feige, "A threshold of $\ln n$ for approximating set cover," *Journal of the ACM*, vol. 45, no. 4, pp. 634–652, 1998.

[28] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997.

[29] R. E. Barlow and F. Proschan, "Importance of system components and fault tree events," *Stochastic Processes and their Applications*, vol. 3, no. 2, pp. 153–173, 1975.

[30] M. A. Harrison, W. L. Ruzzo, and J. D. Ullman, "Protection in operating systems," *Communications of the ACM*, vol. 19, no. 8, pp. 461–471, 1976.

[31] N. Li and M. V. Tripunitara, "Security analysis in role-based access control," *ACM Transactions on Information and System Security*, vol. 9, no. 4, pp. 391–420, 2006.

[32] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture," NIST Special Publication 800-207, National Institute of Standards and Technology, 2020.

[33] T. G. Griffin, F. B. Shepherd, and G. Wilfong, "The stable paths problem and interdomain routing," *IEEE/ACM Transactions on Networking*, vol. 10, no. 2, pp. 232–243, 2002.

Yaroslav Ryabov received the M.S. and Ph.D. degrees in theoretical physics from Kazan State University, Kazan, Russia, in 1993 and 1996, respectively. After research positions at the Hebrew University of Jerusalem, the University of Maryland, Purdue University, and the National Institutes of Health, he moved to applied machine learning and now works as a lead data scientist in financial regulation, building production agent-based systems and evaluation methods for large language models. His current research interests include authorization and trust in multi-agent systems, cooperative game theory, and the computational complexity of access control. He has authored 42 peer-reviewed papers, holds six patents, and received a Bill and Melinda Gates Foundation research award and a National Research Council research associateship.

Supplemental Material for "The Game of Peers: Toward a Credential Algebra for Human-AI Actor Networks"

Yaroslav Ryabov

Abstract—This document accompanies the main text. It contains the full proofs of all results stated there (Section S1, with the main text's numbering), additional results and remarks (S2), the mechanism-design layer (S3), the discussion of the simplifying assumptions (S4), the discussion of the worked examples (S5), the extended open problems (S6), a notation reference (S7), the full development of the worked examples (S8–S9), the exhaustive coalition computation for Example 4 (S10), and a computational evaluation on random instances (S11).

S1 Full Proofs for the Results of the Main Text

Results are restated with the number they carry in the main text; new results introduced only here are numbered S1, S2, ... Section and equation numbers of the form "Section 4" or "(6)" refer to the main text; equation numbers of the form (S1) refer to this document.

S1.1 The Credential Algebra (Section 3)

S1.1.1 Actors, Credentials, and Operators (Section 3.1)

Proposition 3.1 (Properties of Fundamental Operators (restated))**.** *The four fundamental operators satisfy the following:*

Operator	*Commutative*	*Associative*	*Idempotent*	*Monotone*	*Bounded*
$\cap$ *(intersection)*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
$\cup$ *(union)*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
$\ltimes$ *(left join)*	*No*	*No*	*Yes*	*Yes*	*Yes*
$\rtimes$ *(right join)*	*No*	*No*	*Yes*	*Yes*	*Yes*

Proof: The properties of $\cap$ and $\cup$ are standard from set theory. For the left join $S_1 \ltimes S_2 = S_1 \cup (S_2 \cap R)$: commutativity fails since $S_2 \ltimes S_1 = S_2 \cup (S_1 \cap R) \neq S_1 \ltimes S_2$ in general. Associativity fails by the counterexample of Proposition S5 below. Idempotency: $S \ltimes S = S \cup (S \cap R) = S$. Monotonicity and boundedness follow from monotonicity and boundedness of $\cap$ and $\cup$. The right join is symmetric. □

Definition 3.8 (Pass-Filtered Operators (restated))**.** *A credential operator $\otimes$ is* pass-filtered *if there are a set $\pi \subseteq \mathcal{C}$ (the* child filter*) and a monotone map $\alpha : \mathcal{P}(\mathcal{C}) \to \mathcal{P}(\mathcal{C})$ (the* parent filter*) with $\alpha(Y) \supseteq Y \setminus \pi$ for all Y, such that*

$$\otimes(X, Y) \;=\; \big(X \cap \alpha(Y)\big) \,\cup\, (Y \cap \pi) \qquad \text{for all } X, Y \subseteq \mathcal{C}.$$

We write $\mathcal{P}$ for the class of pass-filtered operators. The four fundamental operators are pass-filtered: $\cap$ has $\alpha(Y) = Y$, $\pi = \emptyset$; $\cup$ has $\alpha \equiv \mathcal{C}$, $\pi = \mathcal{C}$; $\ltimes_R$ has $\alpha \equiv \mathcal{C}$, $\pi = R$; and $\rtimes_R$ has $\alpha \equiv R$, $\pi = \mathcal{C}$.

Lemma 3.1 (Properties of Pass-Filtered Operators (restated))**.** *Every $\otimes \in \mathcal{P}$ with filters (α, π) is monotone in both arguments, bounded ($\otimes(X, Y) \subseteq X \cup Y$), and union-distributive in its first argument; its pass filter (Definition 3.7 of the main text) is $\pi(\otimes) = \pi$; and for all X, Y,*

$$Y \cap \pi \;\subseteq\; \otimes(X, Y) \;\subseteq\; X \cup (Y \cap \pi),$$
$$\otimes(X, Y) \cap Y \;=\; Y \cap (X \cup \pi).$$

Proof: Monotonicity in X and boundedness are immediate from the form; monotonicity in Y follows from that of α. Distributivity: $((X_1 \cup X_2) \cap \alpha(Y)) \cup (Y \cap \pi)$ is the union of $(X_i \cap \alpha(Y)) \cup (Y \cap \pi)$ over $i = 1, 2$. The two inclusions are read off the form, with $\otimes(\emptyset, Y) = Y \cap \pi$ showing that no larger set than π passes for every X. Exactness: $\otimes(X, Y) \cap Y = (X \cap \alpha(Y) \cap Y) \cup (Y \cap \pi)$; the inclusion $\supseteq Y \cap (X \cup \pi)$ holds because a credential $c \in X \cap Y$ with $c \notin \pi$ lies in $Y \setminus \pi \subseteq \alpha(Y)$, and $\subseteq$ is clear. □

S1.1.2 Propagation and Coalition Effective Sets (Section 3.2)

Definition 3.11 (Propagation Equations on a Directed Graph (restated))**.** *Let $\mathcal{G} = (\mathcal{A}, E)$ be a directed graph without self-loops, and let $O \subseteq \mathcal{A}$ be a set of* origins*: the actors at which a query starts (on a rooted tree, $O = \{a_0\}$). For $S \subseteq \mathcal{A}$ let $F_S = (S, E_S)$ be the subgraph induced by S; the* parents *of a in F_S are $\mathrm{par}_S(a) \coloneqq \{p \in S : (p, a) \in E\}$, the* sources *of F_S are the actors without parents, $\mathrm{src}(F_S) \coloneqq \{a \in S : \mathrm{par}_S(a) = \emptyset\}$, and*

$$O_S \coloneqq (O \cap S) \cup \mathrm{src}(F_S)$$

for the origins of the coalition: designated origins that belong to S, together with the actors that S cuts off from every parent and that therefore revert to their intrinsic sets, exactly as component roots do in Definition 3.11 of the main text. A credential assignment

on S is a map $E : S \to \mathcal{P}(\mathcal{C})$; assignments are ordered pointwise by inclusion, so that $\mathcal{P}(\mathcal{C})^S$ is a finite complete lattice with least element $\bot$ (every set empty). The propagation map $\Theta_S : \mathcal{P}(\mathcal{C})^S \to \mathcal{P}(\mathcal{C})^S$ *is*

$$\Theta_S(E)(a) \coloneqq \big[\sigma(a)\big]_{a\in O_S} \cup \bigcup_{p\in \mathrm{par}_S(a)} \otimes_{(p,a)}(E(p),\, \sigma(a))\,, \tag{S1}$$

where $[\sigma(a)]_{a\in O_S}$ denotes $\sigma(a)$ if $a \in O_S$ and $\emptyset$ otherwise. A solution *is an assignment with $\Theta_S(E) = E$. Under full disclosure, the* effective credential sets within S *are given by the* least *solution, $\hat\sigma_S(a) \coloneqq (\mathrm{lfp}\,\Theta_S)(a)$ (which exists by Proposition 3.2 below), and the coalition effective set is $\hat\sigma_S \coloneqq \bigcup_{a\in S}\hat\sigma_S(a)$ as before. On a rooted tree the root is the unique source of every component containing it, so origins beyond the sources are immaterial there; we take $O = \emptyset$ unless stated otherwise. Equation* (S1) *aggregates parents* disjunctively*: a credential reaches a through any parent whose edge passes it, in the spirit of disjunctive permission structures; on trees, where every actor has at most one parent, nothing is aggregated. The conjunctive alternative—replacing the union over parents by an intersection—is a different modeling choice, treated in Definition S2 and Propositions S6–S7: the dichotomy is insensitive to it under shared-outcome evaluation, but at a target it can destroy value by itself.*

Proposition 3.2 (Least Solution and Evaluation (restated))**.** *For every directed graph, operator assignment, coalition S, and origin set O:*

(i) Θ_S is monotone, and its solutions form a nonempty complete lattice under pointwise inclusion. The least solution is the limit of the increasing chain $\bot \subseteq \Theta_S(\bot) \subseteq \Theta_S^2(\bot) \subseteq \cdots$, which is stationary after at most $|S|\cdot|\mathcal{C}|$ steps of cost $O(|E_S|\cdot|\mathcal{C}|)$ each; hence $\hat\sigma_S$, $v(S) = |\hat\sigma_S|$, and every marginal contribution are computable in polynomial time.

(ii) If F_S is acyclic, the solution is unique; it is computed by one pass in topological order in time $O((|S| + |E_S|)\cdot|\mathcal{C}|)$, and on a tree it coincides with Definition 3.11 of the main text.

(iii) Consequently, NP-membership of DESTRUCTIVE-ACTOR *and the sampling bound of Proposition 6.4 hold verbatim on arbitrary directed graphs.*

Proof: (i) Each fundamental operator is monotone in its first argument (Proposition 3.1), union is monotone, and the base term $[\sigma(a)]_{a\in O_S}$ does not depend on E; hence Θ_S is monotone, and the Knaster–Tarski theorem [S1] yields the lattice of solutions. By monotonicity, $\Theta_S^k(\bot)$ is an increasing chain; each non-stationary step adds at least one credential to at least one actor, so the chain is stationary within $|S|\cdot|\mathcal{C}|$ steps, and its limit is a solution below every solution, i.e., the least one. A step performs one operator application per edge. (ii) On an acyclic F_S the equation for a involves only the parents of a, so induction along a topological order determines E uniquely and computes it in one pass; on a tree every actor has at most one parent, $O_S = \mathrm{src}(F_S)$ is the set of component roots, and (S1) is (4 of the main text). (iii) A coalition is a polynomial-size certificate whose value is computed in polynomial time by (i), and the Hoeffding argument of Proposition 6.4 requires only a polynomial-time value oracle. □

Proposition 3.3 (Delegation as Intersection on a Tree (restated))**.** *Let $\mathcal{G}$ be a directed tree rooted at a human actor $h \in \mathcal{A}_H$, with all edges equipped with the intersection operator $\cap$. Write $a_0 = h$ and let $h = a_0 \to a_1 \to \cdots \to a_n$ be the path to an actor a_n at depth n. Then:*

$$\hat\sigma(a_n) = \sigma(h) \cap \sigma(a_1) \cap \cdots \cap \sigma(a_n) \subseteq \sigma(h) \tag{S2}$$

That is, the effective credential set at any point in the tree is a subset of the root's credentials. This recovers the classical property that delegation can only narrow permissions.

Proof: By induction on depth. Base case: $\hat\sigma(a_0) = \sigma(h)$. Inductive step: $\hat\sigma(a_k) = \hat\sigma(a_{k-1}) \cap \sigma(a_k) \subseteq \hat\sigma(a_{k-1}) \subseteq \sigma(h)$, where the last inclusion is the inductive hypothesis. Since intersection is associative and commutative, the order of composition does not matter and the result is $\bigcap_{i=0}^{n}\sigma(a_i)$. □

S1.2 Structural Theory (Section 4)

S1.2.1 Monotonicity and Destructive Intermediaries (Section 4.1)

Theorem 4.1 (Monotonicity Dichotomy (restated))**.** *Let $\mathcal{G}$ be a rooted tree with operators drawn from $\mathcal{P}$ (in particular, from the four fundamental operators). The following are equivalent:*

(i) every edge of $\mathcal{G}$ is preserving, i.e., $\pi(\otimes) = \mathcal{C}$ (among the fundamental operators: $\cup$, $\rtimes_R$ for any R, or $\ltimes_{\mathcal{C}}$);

(ii) for every credential assignment σ and all coalitions $S \subseteq T$, we have $\hat\sigma_S \subseteq \hat\sigma_T$—hence $v(S) \le v(T)$ for every monotone W.

Moreover, if (i) fails on some edge, there exists a credential assignment with all intrinsic sets nonempty and coalitions $S \subset T$ with $v(T) < v(S)$ under $W = |\cdot|$.

Proof: (i) ⇒ (ii). It suffices to treat $T = S \cup \{x\}$ and iterate. Compare the recursion (4 of the main text) on F_S and F_T. Actors outside the subtree of x are unaffected: propagation flows only from parent to child, so every actor whose path from its component root does not pass through x has $\hat\sigma_T(a) = \hat\sigma_S(a)$; this covers, in particular, x's parent and everything above and beside it. The actor x itself contributes $\hat\sigma_T(x)$, a pure addition. Finally, each child c of x in S was a component root of F_S (its parent x was absent), contributing $\hat\sigma_S(c) = \sigma(c)$; in F_T it becomes a child, contributing $\hat\sigma_T(c) = \otimes_{(x,c)}(\hat\sigma_T(x), \sigma(c)) \supseteq \sigma(c) \cap \pi = \sigma(c)$, by the lower pass bound of Lemma 3.1 of the main text and $\pi = \mathcal{C}$. Descendants of c then have their first-argument inputs weakly enlarged while their own second arguments are unchanged;

two properties of the class $\mathcal{P}$ are used here, and both are supplied by Lemma 3.1: the lower pass bound just applied, and monotonicity of every $\otimes \in \mathcal{P}$ in its first argument ($X \mapsto (X \cap \alpha(Y)) \cup (Y \cap \pi)$ is monotone because intersection and union are). Induction down each subtree gives $\hat{\sigma}_T(a) \supseteq \hat{\sigma}_S(a)$ for all $a \in S$. Hence $\hat{\sigma}_T \supseteq \hat{\sigma}_S$.

Converse (gadget construction). Suppose edge (a, b) carries a filtering operator, i.e., $\pi \neq \mathcal{C}$; fix a credential $z \notin \pi$ and any credential $w \neq z$, with $w \in \pi$ if $\pi \neq \emptyset$. Assign $\sigma(a) = \{w\}$ and $\sigma(b) = \{z, w\}$ (all other actors may hold arbitrary nonempty sets disjoint from $\{z, w\}$; they are not used). Take $S = \{b\}$ and $T = \{a, b\}$. Then $\hat{\sigma}_S = \{z, w\}$, so $v(S) = 2$, while by Lemma 3.1 of the main text (exactness), $z \notin \hat{\sigma}_T(b) \subseteq \{w\} \cup (\{z, w\} \cap \pi)$ and $\hat{\sigma}_T = \{w\} \cup \hat{\sigma}_T(b) = \{w\}$, so $v(T) = 1 < 2$. The gadget is local to the edge (a, b), so no topological restriction on the remainder of $\mathcal{G}$ can restore monotonicity. □

Proposition 4.1 (Credential protection (restated))**.** *Let $\mathcal{G}$ be a rooted tree with operators in $\mathcal{P}$, and fix the assignment σ.*

(i) If $h \in S$ is a c-protected holder of c, then $c \in \hat{\sigma}_S(h)$; hence $c \in \hat{\sigma}_T$ for every coalition $T \ni h$.

(ii) $\hat{\sigma}_S \subseteq \hat{\sigma}_T$ for all $S \subseteq T$—no credential is ever lost, and v is monotone for every monotone W—if and only if every credential is protected.

(iii) Whether every credential is protected is decided in $O(|\mathcal{A}| \cdot |\mathcal{C}|)$ time by one pass from the root.

Proof: (i) By induction on the depth of h. If the parent p of h is not in S (in particular if h is the root), h is a component root of F_S and $\hat{\sigma}_S(h) = \sigma(h) \ni c$. Otherwise $\hat{\sigma}_S(h) = \otimes_{(p,h)}(\hat{\sigma}_S(p), \sigma(h))$, and by the exactness clause of Lemma 3.1 of the main text, $c \in \hat{\sigma}_S(h)$ if and only if $c \in \hat{\sigma}_S(p) \cup \pi(\otimes_{(p,h)})$. If $c \in \pi$ we are done; otherwise $c \in \sigma(p)$ and p is c-protected with smaller depth, so $c \in \hat{\sigma}_S(p)$ by induction. The second claim follows because $h \in T$.

(ii) *If.* Let $c \in \hat{\sigma}_S$, say $c \in \hat{\sigma}_S(a)$. By the upper pass bound, $\hat{\sigma}_S(a) \subseteq \hat{\sigma}_S(p) \cup \sigma(a)$ for the parent p of a when $p \in S$, so by induction along the path from a's component root, $c \in \sigma(h)$ for some $h \in S$ on that path. By hypothesis h is c-protected, and (i) gives $c \in \hat{\sigma}_T$ for every $T \supseteq S$. *Only if.* Let h be a holder of c that is not c-protected, and let $h = h_0, h_1, \ldots, h_k$ be the maximal ascending run of holders of c (h_{i+1} the parent of h_i). By Definition 4.1 of the main text the run does not reach the root, so h_k has a parent x with $c \notin \sigma(x)$, and no edge (h_{i+1}, h_i) or (x, h_k) passes c. Take $S = \{h_0, \ldots, h_k\}$ and $T = S \cup \{x\}$. In F_S, h_k is a component root, so $c \in \hat{\sigma}_S$. In F_T, x is a component root with $\hat{\sigma}_T(x) = \sigma(x) \not\ni c$; descending the run, exactness gives $c \notin \hat{\sigma}_T(h_i)$ at each step, since $c \notin \hat{\sigma}_T(h_{i+1}) \cup \pi$ (with $h_{k+1} \coloneqq x$). Hence $c \notin \hat{\sigma}_T$, and $W = \mathbf{1}[c \in \cdot\,]$ is a monotone valuation with $v(T) = 0 < 1 = v(S)$.

(iii) Process actors in any top-down order. The root is protected for every credential; for h with parent p, h is c-protected if and only if $c \in \pi(\otimes_{(p,h)})$ or ($c \in \sigma(p)$ and p is c-protected), a constant-time lookup per pair (h, c) once p has been processed. □

Proposition 4.2 (Leaf joiners are harmless (restated))**.** *Let $\mathcal{G}$ be a rooted tree. If x has no children in S, then $\Delta_x(S) = v(S \cup \{x\}) - v(S) \geq 0$, regardless of the operator on the edge connecting x to its parent—including intersection.*

Proof: With no children of x in S, no actor of S changes component or input: $\hat{\sigma}_T(a) = \hat{\sigma}_S(a)$ for all $a \in S$, and $\hat{\sigma}_T = \hat{\sigma}_S \cup \hat{\sigma}_T(x) \supseteq \hat{\sigma}_S$. Monotone W gives $\Delta_x(S) \geq 0$. □

Theorem 4.2 (Destructive Intermediary Criterion (restated))**.** *Let $\mathcal{G}$ be a rooted tree, $x \notin S$, $T = S \cup \{x\}$, and $W = |\cdot|$. Let $b_1, \ldots, b_m$ be the children of x in S whose edges are filtering, with pass filters $\pi_1, \ldots, \pi_m$, and suppose that each b_i is either isolated in F_S or connected downward in F_S only through union edges; the remaining children of x in S attach through preserving edges and may carry arbitrary subtrees. Let D be the set of members of S below some b_i, write*

$$\hat{\sigma}_T(x) = \begin{cases} \otimes_{(p,x)}(\hat{\sigma}_S(p), \sigma(x)), & \text{if } x\text{'s parent } p \in S, \\ \sigma(x), & \text{otherwise}, \end{cases}$$

and define the independent coverage *and the* loss set

$$N \coloneqq \bigcup_{a \in S \setminus (\{b_1, \ldots, b_m\} \cup D)} \hat{\sigma}_S(a) \;\cup\; \bigcup_{d \in D} \sigma(d),$$

$$L \coloneqq \Big(\bigcup_{i=1}^{m} \sigma(b_i)\Big) \setminus \Big(N \cup \hat{\sigma}_T(x) \cup \bigcup_{i=1}^{m} (\sigma(b_i) \cap \pi_i)\Big).$$

Then

$$\Delta_x(S) = \big|\hat{\sigma}_T(x) \setminus \hat{\sigma}_S\big| \;-\; |L|, \tag{S3}$$

and L is exactly the set of credentials that x's presence destroys. In particular, x is destructive ($\Delta_x(S) < 0$) if and only if the credentials held only by its filtered children that neither survive their filters nor are resupplied by x outnumber x's genuinely novel contribution. When the filtered children are isolated in F_S and no credential is held by two of them without being held elsewhere in S, the loss set splits by child: with $U_i \coloneqq \sigma(b_i) \setminus \bigcup_{a \in S \setminus \{b_i\}} \hat{\sigma}_S(a)$ the credentials covered in $\hat{\sigma}_S$ only by b_i,

$$|L| = \sum_{i=1}^{m} \big|U_i \setminus (\hat{\sigma}_T(x) \cup \pi_i)\big|. \tag{S4}$$

Proof: Compare $\hat{\sigma}_S$ with $\hat{\sigma}_T$. Members of S that are not below x keep their sets. A preserving child of x was a component root contributing σ and now receives $\otimes(\hat{\sigma}_T(x), \sigma) \supseteq \sigma$ by the lower pass bound; by monotonicity in the first

argument the same holds along its subtree, and by the upper pass bound every added credential lies in $\hat{\sigma}_T(x)$. A filtered child b_i was a component root with $\hat{\sigma}_S(b_i) = \sigma(b_i)$ and now receives $\otimes_{(x,b_i)}(\hat{\sigma}_T(x), \sigma(b_i))$, which by the exactness clause of Lemma 3.1 of the main text retains $z \in \sigma(b_i)$ if and only if $z \in \hat{\sigma}_T(x) \cup \pi_i$ and contains nothing outside $\hat{\sigma}_T(x) \cup \sigma(b_i)$. A descendant $d \in D$, attached through union edges, holds $\hat{\sigma}(\text{parent}) \cup \sigma(d)$ in both F_S and F_T: its independent contribution is $\sigma(d)$, and everything else it holds is inherited from b_i's set. Hence $\hat{\sigma}_T \setminus \hat{\sigma}_S \subseteq \hat{\sigma}_T(x)$, and every credential of $\hat{\sigma}_T(x) \setminus \hat{\sigma}_S$ is present in $\hat{\sigma}_T$ (in $\hat{\sigma}_T(x)$ itself); the gain is $|\hat{\sigma}_T(x) \setminus \hat{\sigma}_S|$. Conversely, a credential $z \in \hat{\sigma}_S$ is absent from $\hat{\sigma}_T$ if and only if every member that held it was a filtered child or one of its descendants inheriting it (so $z \notin N$), no filtered child retains it ($z \notin \hat{\sigma}_T(x)$ and $z \notin \sigma(b_i) \cap \pi_i$ for all i), and nothing new supplies it ($z \notin \hat{\sigma}_T(x)$)—that is, if and only if $z \in L$. Gains lie in $\hat{\sigma}_T(x) \setminus \hat{\sigma}_S$ and losses in $L \subseteq \hat{\sigma}_S \setminus \hat{\sigma}_T(x)$, so $|\hat{\sigma}_T| - |\hat{\sigma}_S|$ is their difference. For the per-child form: with isolated filtered children, $N = \bigcup_{a \in S \setminus \{b_1, \ldots, b_m\}} \hat{\sigma}_S(a)$, and under the proviso each $z \in L$ is held by exactly one filtered child b_i, so $z \in U_i$; conversely $U_i \setminus (\hat{\sigma}_T(x) \cup \pi_i) \subseteq L$. The sets U_i are pairwise disjoint by definition, which gives (S4). Without the proviso a credential held by two filtered children and by nothing else is destroyed but belongs to no U_i, which is why (S3) is stated through L. □

S1.2.2 Beyond Trees: Lifting the Structural Results (Section 4.3)

Lemma S1 (Union-distributivity in the first argument)**.** *Each pass-filtered operator, in particular each fundamental operator, satisfies $\otimes(X_1 \cup X_2,\, Y) = \otimes(X_1, Y) \cup \otimes(X_2, Y)$ for all $X_1, X_2, Y \subseteq \mathcal{C}$; consequently $\otimes(\bigcup_i X_i,\, Y) = \bigcup_i \otimes(X_i, Y)$ for every finite* nonempty *family.*

Proof: For the class, this is Lemma 3.1; for the four instances, four one-line verifications. Intersection: $(X_1 \cup X_2) \cap Y = (X_1 \cap Y) \cup (X_2 \cap Y)$ is distributivity of $\cap$ over $\cup$. Union: both sides equal $X_1 \cup X_2 \cup Y$. Left join: both sides equal $X_1 \cup X_2 \cup (Y \cap R)$. Right join: $Y \cup ((X_1 \cup X_2) \cap R) = Y \cup (X_1 \cap R) \cup (X_2 \cap R)$, which is also the value of $(Y \cup (X_1 \cap R)) \cup (Y \cup (X_2 \cap R))$. The finite-family form follows by induction. (Nonemptiness matters: $\otimes(\emptyset, Y) = Y \cap \pi(\otimes)$ need not be empty, so the identity fails for the empty family.) □

Theorem 4.3 (Monotonicity Dichotomy on Directed Graphs (restated))**.** *Let $\mathcal{G}$ be any directed graph with operators drawn from $\mathcal{P}$ and origin set O, under the semantics of Definition 3.11. The following are equivalent:*

(i) every edge (p, a) whose head a is not an origin ($a \notin O$) is preserving;
(ii) for every credential assignment σ and all coalitions $S \subseteq T$: $\hat{\sigma}_S \subseteq \hat{\sigma}_T$.

Moreover, if (i) fails on some edge, there exist a credential assignment with all intrinsic sets nonempty and coalitions $S \subset T$ with $v(T) < v(S)$ under $W = |\cdot|$. For $O = \emptyset$, and on rooted trees for any $O \subseteq \{a_0\}$, condition (i) reads: every edge is preserving.

Proof: (i) ⇒ (ii). We show $\hat{\sigma}_U = \bigcup_{a \in U} \sigma(a)$ for every coalition U, which is monotone in U. Upper bound: by the upper bound of Lemma 3.1 of the main text and induction on k, $\Theta^k_U(\bot)(a) \subseteq \bigcup_{b \in U} \sigma(b)$. Lower bound: if $a \in O_U$ the base term gives $\Theta_U(\bot)(a) \supseteq \sigma(a)$; otherwise $a \notin O$ and $a \notin \mathrm{src}(F_U)$, so a has a parent p in F_U and the edge (p, a) is preserving by (i), whence $\Theta_U(\bot)(a) \supseteq \otimes_{(p,a)}(\emptyset, \sigma(a)) \supseteq \sigma(a) \cap \mathcal{C} = \sigma(a)$ by Lemma 3.1 of the main text. Hence $\hat{\sigma}_U(a) \supseteq \sigma(a)$ for all $a \in U$.

Converse. Let (a, b) be filtering with $b \notin O$ and $z \notin \pi$. Take $\sigma(a) = \{w\}$, $\sigma(b) = \{z, w\}$ with $w \in \pi$ if $\pi \neq \emptyset$, $S = \{b\}$, $T = \{a, b\}$. In F_S the actor b is a source, so $\hat{\sigma}_S = \{z, w\}$ and $v(S) = 2$. In F_T, b is neither an origin nor a source, so no base term contains z; and z never enters any iterate: by the upper bound of Lemma 3.1 of the main text, $\otimes(X, \sigma(v)) \subseteq X \cup (\sigma(v) \cap \pi_\otimes)$, where $\sigma(a) \not\ni z$ and $\sigma(b) \cap \pi \not\ni z$ for the edge into b; the only other possible edge, (b, a), has second argument $\sigma(a) \not\ni z$. By induction on k, $z \notin \Theta^k_T(\bot)(a) \cup \Theta^k_T(\bot)(b)$, so $\hat{\sigma}_T \subseteq \{w\}$ and $v(T) \leq 1 < 2 = v(S)$. □

Lemma S2 (Where a credential enters)**.** *Let $\mathcal{G}$ be a directed graph with operators in $\mathcal{P}$, origins O, and coalition T. Then $c \in \hat{\sigma}_T$ if and only if some holder $u \in T$ of c is an origin of T ($u \in O_T$) or has a parent $p \in T$ whose edge passes c ($c \in \pi(\otimes_{(p,u)})$).*

Proof: If $u \in O_T$ then $c \in \sigma(u) \subseteq \Theta_T(\bot)(u)$; if $p \in \mathrm{par}_T(u)$ passes c then $c \in \sigma(u) \cap \pi \subseteq \otimes_{(p,u)}(\emptyset, \sigma(u)) \subseteq \Theta_T(\bot)(u)$; either way $c \in \hat{\sigma}_T$. Conversely, suppose no holder in T is an origin of T or has a passing parent in T. By the upper pass bound, $c \in \otimes_{(p,a)}(X, \sigma(a))$ requires $c \in X$ or $c \in \sigma(a) \cap \pi(\otimes_{(p,a)})$; the second alternative is excluded on every edge of F_T, and the base terms contain c only at holders in O_T, of which there are none. Induction on k gives $c \notin \Theta^k_T(\bot)(a)$ for all a and k. □

Definition S1 (Protected holders on a directed graph)**.** *Let $\mathcal{G}$ be a directed graph with operators in $\mathcal{P}$ and origins O, and let $c \in \sigma(h)$. Call h c-*protected *if $h \in O$, or if for every parent p of h in $\mathcal{G}$ either $c \in \pi(\otimes_{(p,h)})$ or ($c \in \sigma(p)$ and p is c-protected); an actor without parents is protected vacuously. On a cyclic graph the recursion is read as its least fixed point (iterate from "no actor is protected"), so that a cycle of holders that filter c from one another protects none of them. On a rooted tree this is Definition 4.1 of the main text.*

Proposition S1 (Credential protection beyond trees)**.** *Let $\mathcal{G}$ be a directed graph with operators in $\mathcal{P}$ and origins O, and fix σ.*

(i) (Sufficiency, all graphs.) *If $h \in S$ is a c-protected holder, then $c \in \hat{\sigma}_S(h)$. If every holder of every credential is protected, then $\hat{\sigma}_S \subseteq \hat{\sigma}_T$ for all $S \subseteq T$.*
(ii) (Exactness on acyclic graphs.) *If $\mathcal{G}$ is acyclic, the converse holds: $\hat{\sigma}_S \subseteq \hat{\sigma}_T$ for all $S \subseteq T$ if and only if every holder is protected; the test takes $O((|\mathcal{A}| + |E|) \cdot |\mathcal{C}|)$ time in topological order.*

(iii) (Failure of exactness on cycles.) *On the three-cycle $u_1 \to u_3 \to u_2 \to u_1$ with intersection on all three edges, an additional union edge $u_1 \to u_2$, $O = \emptyset$, and $\sigma(u_i) = \{c\}$ for all i, no holder is protected, yet $\hat{\sigma}_S \subseteq \hat{\sigma}_T$ for all $S \subseteq T$.*

Proof: (i) Induction on the stage at which h becomes protected in the least-fixed-point iteration. If $h \in O \cap S$ or $\mathrm{par}_S(h) = \emptyset$, then $h \in O_S$ and the base term gives $c \in \hat{\sigma}_S(h)$. Otherwise pick any $p \in \mathrm{par}_S(h)$: if $c \in \pi(\otimes_{(p,h)})$ then $c \in \sigma(h) \cap \pi \subseteq \otimes_{(p,h)}(\hat{\sigma}_S(p), \sigma(h)) \subseteq \hat{\sigma}_S(h)$; else $c \in \sigma(p)$ and p became protected at an earlier stage, so $c \in \hat{\sigma}_S(p)$ by induction and $c \in \otimes_{(p,h)}(\hat{\sigma}_S(p), \sigma(h))$ by exactness. For the second claim, if $c \in \hat{\sigma}_S$ then $c \in \sigma(h)$ for some $h \in S$ (the upper pass bound gives $\hat{\sigma}_S \subseteq \bigcup_{a \in S} \sigma(a)$, as in the proof of Theorem 4.3); h is protected and remains in T, so $c \in \hat{\sigma}_T$.

(ii) Let h be an unprotected holder of c. On an acyclic graph the recursion of Definition S1 is well founded, so $h \notin O$ and some parent p of h has $c \notin \pi(\otimes_{(p,h)})$ and either $c \notin \sigma(p)$ or p unprotected. Iterating, and using acyclicity to rule out repetition, we obtain a path $h = u_0, u_1, \ldots, u_k, x$ in which $u_0, \ldots, u_k$ are unprotected holders, x is not a holder, and every edge (u_{i+1}, u_i) and (x, u_k) filters c. Take $S = \{u_k\}$ and $T = \{u_k, x\}$. Then $c \in \hat{\sigma}_S$, while in F_T the only holder u_k is not an origin, is not a source, and has the single parent x, whose edge filters c; by Lemma S2, $c \notin \hat{\sigma}_T$. For the running time, process actors in topological order; each pair (h, c) is decided by one inspection of the parents of h.

(iii) By Lemma S2, $c \notin \hat{\sigma}_T$ requires every $u_i \in T$ to have a parent in T and no passing parent in T. If $u_1 \in T$, its only parent u_2 is in T; then u_2 has the passing parent u_1 in T, a contradiction. If $u_2 \in T$ and $u_1 \notin T$, the only remaining parent u_3 is in T, whose only parent is u_1—a contradiction; and $u_3 \in T$ forces $u_1 \in T$ directly. So $c \in \hat{\sigma}_T$ for every nonempty T, and no credential is ever lost. As for protection: with $O = \emptyset$ the iteration from "none protected" promotes u_1 only after u_2, u_2 only after u_3, and u_3 only after u_1; nothing is ever promoted. □

Proposition 4.3 (Monotonicity of a deployment is coNP-complete on cyclic graphs (restated))**.** MONOTONE-DEPLOYMENT*—given $\mathcal{G}$, its operators, origins O, and σ, is $\hat{\sigma}_S \subseteq \hat{\sigma}_T$ for all $S \subseteq T$?—is decidable in $O((|\mathcal{A}| + |E|) \cdot |\mathcal{C}|)$ time on acyclic graphs (Proposition S1(ii)) and coNP-complete on directed graphs with cycles, already with $|\mathcal{C}| = 1$, $O = \emptyset$, operators $\cap$ and $\cup$ only, and every actor holding the credential. In the same setting* DESTRUCTIVE-ACTOR *is NP-complete; hence the fixed-parameter tractability in $|\mathcal{C}|$ of Proposition 6.3 does not extend to cyclic graphs unless* $\mathrm{P} = \mathrm{NP}$.

Proof: Membership. If $\hat{\sigma}_S \not\subseteq \hat{\sigma}_T$ for some $S \subset T$, adding the actors of $T \setminus S$ one at a time exhibits a single step $S' \subset S' \cup \{x\}$ at which some credential is lost; the pair (S', x) is a polynomial-size certificate of non-monotonicity, checked in polynomial time by Proposition 3.2.

Hardness. We reduce SAT to the complement. Let $\varphi = C_1 \wedge \cdots \wedge C_m$ over variables $x_1, \ldots, x_n$, no clause containing a variable together with its negation. Build $\mathcal{G}$ with a single credential c held by every actor and $O = \emptyset$: a *clause actor* u_j for each clause and a *literal actor* $\ell_{j,k}$ for the k-th literal of C_j. Edges: $\ell_{j,k} \to u_j$ and $u_{j+1} \to \ell_{j,k}$ (indices modulo m), all carrying $\cap$; and, for every pair of literal actors carrying complementary literals, one edge between them carrying $\cup$, oriented from the lexicographically later pair (j, k) to the earlier. Call a coalition T *killing* if $c \notin \hat{\sigma}_T$; by Lemma S2, T is killing if and only if every actor of T has a parent in T and none has a $\cup$-parent in T. Since $c \in \hat{\sigma}_{\{u\}}$ for every single actor u, the deployment is non-monotone if and only if a nonempty killing coalition exists.

If φ is satisfiable, take T to consist of all clause actors and, for each clause, one literal actor whose literal is true. Every u_j has its chosen literal actor as a $\cap$-parent in T, every chosen $\ell_{j,k}$ has u_{j+1} as a $\cap$-parent in T, and no two chosen literal actors are complementary, so no $\cup$-edge lies inside T: T is killing. Conversely, let T be killing. Every actor of T has a $\cap$-parent in T, so from any actor of T one can walk backwards along $\cap$-edges forever inside T; the $\cap$-edges form the necklace $u_1 \leftarrow \ell_{1,\cdot} \leftarrow u_2 \leftarrow \cdots \leftarrow u_m \leftarrow \ell_{m,\cdot} \leftarrow u_1$, whose only closed walks go around it, so T contains every clause actor and at least one literal actor per clause. No two literal actors in T are complementary, since the $\cup$-edge between them would give one of them a passing parent in T. Setting the literals of T true, and the remaining variables arbitrarily, satisfies φ.

Destructive actor. Every killing coalition contains u_1; with $S \coloneqq T \setminus \{u_1\}$ the actors $\ell_{m,\cdot}$ become sources of F_S, so $c \in \hat{\sigma}_S$ and $\Delta_{u_1}(S) = -1$ under $W = |\cdot|$. Hence u_1 is destructive in some coalition if and only if φ is satisfiable, with $|\mathcal{C}| = 1$; membership in NP is Proposition 3.2(iii). □

Proposition 4.4 (Sinks Are Harmless (restated))**.** *Let $x \notin S$ have no edge into S, so that x is a sink of $F_{S \cup \{x\}}$. Then $\hat{\sigma}_{S \cup \{x\}} \supseteq \hat{\sigma}_S$ for every credential assignment, origin set, and operator assignment; equivalently, $\Delta_x(S) \geq 0$ under shared-outcome evaluation with monotone W. This extends Proposition 4.2 to arbitrary directed graphs.*

Proof: No actor of S has x as a parent, so $\mathrm{par}_{S \cup \{x\}}(a) = \mathrm{par}_S(a)$ and $O_{S \cup \{x\}} \cap S = O_S$. Hence the S-coordinates of $\Theta_{S \cup \{x\}}$ form a closed subsystem identical to Θ_S, and the least solution restricted to S is $\hat{\sigma}_S$; therefore $\hat{\sigma}_{S \cup \{x\}} = \hat{\sigma}_S \cup \hat{\sigma}_{S \cup \{x\}}(x)$. □

Proposition 4.5 (When the Origin Matters (restated))**.** *Fix $\mathcal{G}$ and its operators, and consider the grand coalition $S = \mathcal{A}$. The least solution is the same for every origin set $O \supseteq \mathrm{src}(\mathcal{G})$ and every credential assignment σ if and only if every non-source actor a satisfies the* incoming pass cover *condition*

$$\bigcup_{p \in \mathrm{par}(a)} \pi\big(\otimes_{(p,a)}\big) = \mathcal{C}.$$

In particular, origins are immaterial when every edge is preserving, and matter as soon as some actor has, for some credential, only filtering incoming edges.

Proof: If the cover condition holds, then for every assignment E and every non-source a, $\Theta(E)(a) \supseteq \bigcup_p \otimes_{(p,a)}(\emptyset, \sigma(a)) = \bigcup_p \sigma(a) \cap \pi_{(p,a)} = \sigma(a)$ by Lemma 3.1 of the main text, so adding $\sigma(a)$ as a base term changes nothing: the propagation maps for all $O \supseteq \mathrm{src}(\mathcal{G})$ coincide. Conversely, let a be a non-source with $z \notin \bigcup_p \pi_{(p,a)}$; take $\sigma(a) = \{z\}$ and $\sigma(v) = \emptyset$ for $v \neq a$. With $a \in O$, $z \in \hat{\sigma}(a)$. With $O = \mathrm{src}(\mathcal{G})$ no base term contains z, and by the upper bound of Lemma 3.1 of the main text every term $\otimes_{(p,v)}(X, \sigma(v))$ is contained in $X \cup (\sigma(v) \cap \pi_{(p,v)})$, which never contains z unless X does; by induction z appears in no iterate. □

S1.3 Strategic Layer (Section 5)

S1.3.1 The Disclosure Game and Its Equilibria (Section 5.1)

Proposition 5.1 (Non-existence of pure-strategy equilibria (restated))**.** *There exist disclosure games $\Gamma_{a,b}$—with the union operator, monotone value functions, and additive disclosure costs—that admit no pure-strategy Nash equilibrium.*

Proof: We construct an explicit game. Let $\mathcal{C} = \{x, y\}$, with intrinsic credential sets $\sigma(a) = \{x\}$ and $\sigma(b) = \{y\}$, so that the strategy sets are $S_a = \{\emptyset, \{x\}\}$ and $S_b = \{\emptyset, \{y\}\}$. Let the edge operator be union, $\otimes_{a,b}(\delta_{a\to b}, \delta_{b\to a}) = \delta_{a\to b} \cup \delta_{b\to a}$, and let disclosure costs be additive with unit weight: $D_i(\delta) = |\delta|$ for $i \in \{a, b\}$. Define the value functions

$$V_a(T) = \begin{cases} 10 & \text{if } T = \{x, y\} \\ 0 & \text{otherwise,} \end{cases} \qquad V_b(T) = \begin{cases} 0 & \text{if } T = \emptyset \\ 10 & \text{otherwise.} \end{cases}$$

Both value functions are monotone with respect to set inclusion. Semantically, actor a benefits only from the *complete* credential bundle (the credentials are complements for a), while actor b benefits as soon as *any* credential is effective (the credentials are substitutes for b).

Writing strategy profiles as $(\delta_{a\to b}, \delta_{b\to a})$, the four payoff pairs (u_a, u_b) are:

	$\delta_{b\to a} = \emptyset$	$\delta_{b\to a} = \{y\}$
$\delta_{a\to b} = \emptyset$	$(0,\ 0)$	$(0,\ 9)$
$\delta_{a\to b} = \{x\}$	$(-1,\ 10)$	$(9,\ 9)$

We check each profile. At $(\emptyset, \emptyset)$, actor b deviates to $\{y\}$ (payoff $9 > 0$). At $(\emptyset, \{y\})$, actor a deviates to $\{x\}$ (payoff $9 > 0$). At $(\{x\}, \{y\})$, actor b deviates to $\emptyset$ (payoff $10 > 9$: credential x already substitutes, so b saves its disclosure cost). At $(\{x\}, \emptyset)$, actor a deviates to $\emptyset$ (payoff $0 > -1$). No profile is a Nash equilibrium. □

S1.3.2 Equilibrium Classification (Section 5.2)

Proposition 5.2 (No-disclosure deadlock under intersection (restated))**.** *Let the edge operator be intersection, and let the disclosure costs satisfy $D_i(\emptyset) = 0$ and $D_i \geq 0$. Then for* all *value functions V_a, V_b, the profile $(\emptyset, \emptyset)$ is a pure-strategy Nash equilibrium of $\Gamma_{a,b}$. If moreover $D_i \equiv 0$ and each V_i is monotone with respect to set inclusion, full disclosure $(\sigma(a), \sigma(b))$ is also a pure-strategy equilibrium.*

Proof: Fix $\delta_{b\to a} = \emptyset$. For any deviation $\delta_{a\to b}$ we have $\delta_{a\to b} \cap \emptyset = \emptyset$, so the effective set is $\emptyset$ regardless of a's choice, and $u_a(\delta_{a\to b}, \emptyset) = V_a(\emptyset) - D_a(\delta_{a\to b}) \leq V_a(\emptyset) - D_a(\emptyset) = u_a(\emptyset, \emptyset)$. The argument for b is symmetric, so neither actor has a profitable deviation. For the second claim: against $\delta_{b\to a} = \sigma(b)$, any $\delta_{a\to b}$ yields effective set $\delta_{a\to b} \cap \sigma(b) \subseteq \sigma(a) \cap \sigma(b)$, with equality at $\delta_{a\to b} = \sigma(a)$; monotonicity of V_a and costlessness make full disclosure a best response, and symmetrically for b. □

Proposition 5.3 (Modular preferences: equilibria under all four operators (restated))**.** *Let the edge operator be any of the four fundamental operators, and let all V_i and D_i be modular with $V_i(\emptyset) = D_i(\emptyset) = 0$ (hence with nonnegative weights $w_i(c) = V_i(\{c\})$ and $d_i(c) = D_i(\{c\})$). Then $\Gamma_{a,b}$ admits a pure-strategy Nash equilibrium.*

Proof: By Lemma S3, the effectiveness of each credential c depends only on the two disclosure decisions concerning c; by modularity, each actor's utility is the sum over credentials of per-credential terms $w_i(c) \cdot \mathbf{1}[c \text{ effective}] - d_i(c) \cdot \mathbf{1}[c \in \delta_i]$. The disclosure game therefore decomposes into independent 2×2 games, one per credential, and any product of per-credential equilibria is an equilibrium of $\Gamma_{a,b}$. It remains to exhibit an equilibrium of each 2×2 factor game. If c is effective only unilaterally (e.g., $c \in \sigma(a) \setminus \sigma(b)$ under union, or $c \in \sigma(b) \cap R$ with $c \notin \sigma(a)$ under $\ltimes_R$), the holder has a dominant strategy—disclose if and only if $w \geq d$—and the other player's choice concerning c is payoff-irrelevant. If c can never be effective (c unique to one actor under intersection; $c \in \sigma(b) \setminus R$ under $\ltimes_R$), withholding weakly dominates. In the genuinely bilateral cases, effectiveness is a monotone function of both decisions and both players weakly prefer c effective ($w_i(c) \geq 0$), so the factor game is either a coordination game (under $\wedge$: both-withhold is always an equilibrium) or a chicken-type game (under $\vee$: if $w_b(c) \geq d_b(c)$ then (a withholds, b discloses) is an equilibrium; symmetrically for a; if neither inequality holds, mutual withholding is an equilibrium). A matching-pennies pattern—the only 2×2 obstruction to pure equilibria—cannot arise, since it requires the players to rank the outcomes in opposite directions. □

Theorem 5.1 (Strategic complementarity under union (restated))**.** *Let the edge operator be union, let the intrinsic credential sets be disjoint ($\sigma(a) \cap \sigma(b) = \emptyset$), let each V_i be supermodular, and let each D_i be submodular (in particular, modular costs qualify). Then*

$\Gamma_{a,b}$ is a supermodular game on the lattices $(\mathcal{P}(\sigma(a)), \subseteq)$ and $(\mathcal{P}(\sigma(b)), \subseteq)$; consequently it admits a pure-strategy Nash equilibrium, and the equilibrium set contains a largest and a smallest equilibrium [S2].

Proof: We verify the two defining conditions of a supermodular game [S2].

(i) Supermodularity in own strategy. Fix $\delta_{b\to a}$. For $T_1, T_2 \subseteq \sigma(a)$, distributivity of union over intersection gives $(T_1 \cup \delta_{b\to a}) \cup (T_2 \cup \delta_{b\to a}) = (T_1 \cup T_2) \cup \delta_{b\to a}$ and $(T_1 \cup \delta_{b\to a}) \cap (T_2 \cup \delta_{b\to a}) = (T_1 \cap T_2) \cup \delta_{b\to a}$. Applying the supermodular inequality for V_a at the pair of sets $T_1 \cup \delta_{b\to a}$, $T_2 \cup \delta_{b\to a}$ yields supermodularity of $T \mapsto V_a(T \cup \delta_{b\to a})$. Since D_a is submodular, $-D_a$ is supermodular, and $u_a(\cdot, \delta_{b\to a})$ is a sum of supermodular functions, hence supermodular. Symmetrically for b.

(ii) Increasing differences. Let $\delta'_a \subseteq \delta_a \subseteq \sigma(a)$ and $\delta'_b \subseteq \delta_b \subseteq \sigma(b)$. The cost terms cancel from the differences, so it suffices to show

$$V_a(\delta_a \cup \delta_b) - V_a(\delta'_a \cup \delta_b) \;\geq\; V_a(\delta_a \cup \delta'_b) - V_a(\delta'_a \cup \delta'_b).$$

Set $A \coloneqq \delta_a \setminus \delta'_a$, $S_1 \coloneqq \delta'_a \cup \delta'_b$, and $S_2 \coloneqq \delta'_a \cup \delta_b \supseteq S_1$. By disjointness of the intrinsic sets, $A \cap S_2 = \emptyset$. The displayed inequality reads $V_a(S_2 \cup A) - V_a(S_2) \geq V_a(S_1 \cup A) - V_a(S_1)$, which is the standard marginal-value characterization of supermodularity (applied elementwise to A by induction, using the nondecreasing-marginals form of Definition 5.3 of the main text). The argument for u_b is symmetric. By Topkis's theorem [S2], a finite supermodular game admits extremal pure-strategy equilibria. □

Proposition 5.4 (Common interests: the disclosure game is a potential game (restated))**.** *Suppose $V_a = V_b = V$ for an arbitrary value function V, with arbitrary disclosure costs and an* arbitrary *edge operator $\otimes$. Then $\Gamma_{a,b}$ is an exact potential game [S3] with potential*

$$P(\delta_{a\to b}, \delta_{b\to a}) = V(\otimes(\delta_{a\to b}, \delta_{b\to a})) - D_a(\delta_{a\to b}) - D_b(\delta_{b\to a}),$$

and any maximizer of P is a pure-strategy Nash equilibrium; in particular, pure equilibria exist.

Proof: For any unilateral change by a from $\delta'_{a\to b}$ to $\delta_{a\to b}$ (holding $\delta_{b\to a}$ fixed), both u_a and P change by $V(\otimes(\delta_{a\to b}, \delta_{b\to a})) - V(\otimes(\delta'_{a\to b}, \delta_{b\to a})) - D_a(\delta_{a\to b}) + D_a(\delta'_{a\to b})$, since the term $D_b(\delta_{b\to a})$ is constant under the change; symmetrically for b. Thus P is an exact potential. The strategy space is finite, so P attains a maximum, and no unilateral deviation from a maximizer can increase P—hence none can increase the deviator's utility [S3]. □

S1.4 Computational Complexity and Synthesis (Section 6)

S1.4.1 Coalition Value and Attribution (Section 5.3)

Proposition 6.1 (Coalition evaluation is linear on trees (restated))**.** *Given a rooted tree $\mathcal{G}$, an operator assignment, a credential assignment, and a coalition S, the coalition effective set $\hat{\sigma}_S$ and the value $v(S) = |\hat{\sigma}_S|$ can be computed in time $O(|S| \cdot |\mathcal{C}|)$. Consequently, any single marginal contribution $\Delta_x(S)$ is computable in time $O(n \cdot |\mathcal{C}|)$.*

Proof: Traverse each component of F_S from its root downward, applying recursion (4 of the main text); each edge requires one operator application, i.e., $O(|\mathcal{C}|)$ elementary set operations on bit-vectors of length $|\mathcal{C}|$. Accumulate the union and its cardinality on the fly. A marginal contribution is two such evaluations. □

Theorem 6.1 (Detecting destructiveness is NP-complete (restated))**.** *The* DESTRUCTIVE-ACTOR *problem is NP-complete. Hardness holds already for trees of depth one using only the operators $\cap$ and $\cup$, and therefore transfers to every graph class containing such trees; combined with Proposition 3.2, the problem is NP-complete on arbitrary directed graphs as well. Consequently, deciding that a given actor is* never *destructive is coNP-complete.*

Proof: Membership. A coalition S is a certificate: $\Delta_x(S)$ is evaluated in polynomial time by Proposition 6.1.

Hardness. We reduce from RED-BLUE SET COVER [S4]: given disjoint finite sets B (blue) and R (red), a family $Q_1, \ldots, Q_m \subseteq B \cup R$, and a budget k (w.l.o.g. $k \leq |R| - 1$), decide whether some subfamily covers all of B while covering at most k elements of R.

Construct the following instance. The credential universe consists of: $M \coloneqq |R| + 1$ distinct copies $b^{(1)}, \ldots, b^{(M)}$ of each blue element $b \in B$; the red elements R; and $g_0 \coloneqq |R| - k - 1$ fresh *dummy* credentials. The actor set is $\{x, c, q_1, \ldots, q_m\}$, arranged as a depth-one tree rooted at x: the edge (x, c) carries $\cap$, and each edge (x, q_j) carries $\cup$. Credentials: $\sigma(x)$ consists of all blue copies and all dummies; $\sigma(c) = R$; and $\sigma(q_j)$ consists of the blue copies of $Q_j \cap B$ together with $Q_j \cap R$.

Consider any $S \subseteq \{c, q_1, \ldots, q_m\}$; all members of S are isolated in F_S (they are siblings), so Theorem 4.2 applies with $\hat{\sigma}_T(x) = \sigma(x)$ (x is the root). If $c \notin S$, the theorem with $m = 0$ filtered children gives $\Delta_x(S) = |\sigma(x) \setminus \hat{\sigma}_S| \geq 0$. So suppose $S = \{c\} \cup \{q_j : j \in J\}$. Writing $\mathrm{ub}(J)$ for the number of blue elements not covered by $\{Q_j\}_{j\in J}$ and $\mathrm{ur}(J)$ for the number of red elements not covered, we compute: the gain term is $|\sigma(x) \setminus \hat{\sigma}_S| = M \cdot \mathrm{ub}(J) + g_0$ (blue copies are covered exactly when their blue element is; dummies are never covered; $R \cap \sigma(x) = \emptyset$); the unique coverage of the single filtered child is $U = R \setminus \bigcup_{j\in J} Q_j$, and since $\pi(\cap) = \emptyset$ and $U \cap \sigma(x) = \emptyset$, the loss term is $|U| = \mathrm{ur}(J)$. Hence by (S3),

$$\Delta_x(S) = M \cdot \mathrm{ub}(J) + g_0 - \mathrm{ur}(J).$$

If $\mathrm{ub}(J) \geq 1$, then $\Delta_x(S) \geq M = |R| + 1 > \mathrm{ur}(J)$, so x is not destructive. If $\mathrm{ub}(J) = 0$, then $\Delta_x(S) < 0$ if and only if $\mathrm{ur}(J) > g_0 = |R| - k - 1$, i.e., if and only if the subfamily covers at most k red elements. Thus a destructive coalition exists if and only if the RED-BLUE SET COVER instance is a yes-instance. The construction is polynomial. □

Proposition 6.2 (Counting destructive coalitions is #P-complete (restated))**.** *Given an instance as in Definition 6.1 of the main text, computing* $|\{S \subseteq \mathcal{A} \setminus \{x\} : \Delta_x(S) < 0\}|$ *is #P-complete. As with Theorem 6.1, hardness on depth-one trees transfers to every containing graph class.*

Proof: Membership follows since the underlying witness relation is polynomial-time decidable. For hardness, reduce from counting the satisfying assignments of a monotone conjunctive-normal-form (CNF) formula, which is #P-complete [S5]. Given $\varphi = \bigwedge_{i=1}^{n'} \bigvee_{j \in C_i} y_j$ over variables $y_1, \ldots, y_m$, instantiate the construction of Theorem 6.1 with $B = \{b_1, \ldots, b_{n'}\}$ (one blue element per clause, $M = 1$ copy), $Q_j = \{b_i : j \in C_i\}$, $R = \{z\}$ a single red element contained in no Q_j, and $g_0 = 0$ dummies. Then $\mathrm{ur} \equiv 1$, and by the displayed formula, S is destructive if and only if $c \in S$ and the chosen subfamily covers all blue elements—i.e., if and only if the assignment $y_j = 1 \Leftrightarrow q_j \in S$ satisfies every clause. This is a bijection between destructive coalitions and satisfying assignments of φ. □

Proposition 6.3 (The hardness is in the actors, not the credentials (restated))**.** *On rooted trees,* DESTRUCTIVE-ACTOR *is decidable in time* $2^{O(|\mathcal{C}|)} \cdot |\mathcal{A}|$*; the number of coalitions* S *with* $\Delta_x(S) < 0$ *and the exact Shapley value* ϕ_x *are computable in time* $2^{O(|\mathcal{C}|)} \cdot \mathrm{poly}(|\mathcal{A}|)$*. Thus all three problems are fixed-parameter tractable in the number of credentials: the hardness of Theorem 6.1 and Proposition 6.2 lies in the number of actors.*

Proof: Fix x and write $S^+ = S \cup \{x\}$. On a tree, the effective set of an actor $u \in S$ depends only on u's *context*—the effective set of its parent if the parent is in the coalition, and $\perp$ otherwise—and on $\sigma(u)$; children of u receive context $\hat{\sigma}_S(u)$ if $u \in S$ and $\perp$ if not. Hence, given the context at u, the subtrees of u's children are independent. Actors outside the subtree of x have the same context under S and S^+, since propagation flows away from the root only; inside it, the two contexts may differ. For each actor u and pair of contexts $(X^-, X^+) \in (\mathcal{P}(\mathcal{C}) \cup \{\perp\})^2$ let $A_u(X^-, X^+)$ be the set of pairs (Y^-, Y^+) such that, for some choice of $S \cap \mathrm{subtree}(u)$ (with x absent from S and present in S^+), $Y^{\mp}$ is the union of the effective sets of the members of the subtree under S and S^+, respectively. Then A_u is computed bottom-up: for each membership choice at u (forced if $u = x$), the contribution (e^-, e^+) of u itself is determined by the contexts, and A_u is the set of coordinatewise unions of that contribution with one pair from each child's $A_c(e^-, e^+)$ (or $A_c(\perp, \perp)$ if u is absent). Each $A_u(\cdot, \cdot)$ has at most $4^{|\mathcal{C}|}$ elements, there are at most $(2^{|\mathcal{C}|} + 1)^2$ context pairs per actor, and combining a child costs at most $16^{|\mathcal{C}|}$ set operations, so the root value $A_{a_0}(\perp, \perp)$ is obtained in time $2^{O(|\mathcal{C}|)} \cdot |\mathcal{A}|$; x is destructive if and only if it contains a pair with $|Y^+| < |Y^-|$. Replacing sets of pairs by multiplicity tables counts the coalitions realizing each pair, which yields the number of destructive coalitions; recording in addition the coalition size $|S| \leq |\mathcal{A}|$ yields, for each size k, the sum of $|Y^+| - |Y^-|$ over coalitions of size k, from which $\phi_x = \sum_k \frac{k!\,(|\mathcal{A}|-k-1)!}{|\mathcal{A}|!} \sum_{|S|=k} \Delta_x(S)$ follows in polynomial time. □

Proposition 6.4 (Shapley values: exact hardness, efficient approximation (restated))**.** *(i) Exact computation of Shapley values from the definition requires summation over all coalitions; in general succinctly represented cooperative games this is #P-hard [S6]. (ii) In the credential coalition game on a tree, however, the value oracle is polynomial (Proposition 6.1), so the Shapley value admits an efficient randomized approximation: for any* $\varepsilon > 0$ *and* $\delta \in (0, 1)$*, sampling*

$$m \;\geq\; \frac{2|\mathcal{C}|^2 \ln(2/\delta)}{\varepsilon^2}$$

uniformly random permutations and averaging the marginal contributions of x *yields an estimate* $\hat{\phi}_x$ *with* $\Pr\left[|\hat{\phi}_x - \phi_x| \geq \varepsilon\right] \leq \delta$*, in total time* $O(m \cdot n \cdot |\mathcal{C}|)$ *[S7].*

Proof: (ii) The Shapley value is the expectation of the marginal contribution $\Delta_x(\mathrm{pred}_\sigma(x))$ over a uniformly random ordering σ (Definition 5.6 of the main text). With $W = |\cdot|$, every coalition value lies in $[0, |\mathcal{C}|]$, so every marginal contribution lies in $[-|\mathcal{C}|, |\mathcal{C}|]$, an interval of length $2|\mathcal{C}|$. Hoeffding's inequality for m i.i.d. samples bounded in an interval of length $2|\mathcal{C}|$ gives $\Pr[|\hat{\phi}_x - \phi_x| \geq \varepsilon] \leq 2\exp\left(-\frac{m\varepsilon^2}{2|\mathcal{C}|^2}\right)$, which is at most δ for the stated m. Each sample costs one marginal evaluation, i.e., $O(n \cdot |\mathcal{C}|)$ by Proposition 6.1. □

Proposition 6.5 (Preserving regime: exactly weighted set cover (restated))**.** *If every edge of* $\mathcal{G}$ *is preserving, then* $\hat{\sigma}_S = \bigcup_{a \in S} \sigma(a)$ *for every coalition* S*, and* MCAC *is equivalent to weighted* SET COVER *with universe* Q*, sets* $\sigma(a) \cap Q$*, and weights* $c(a)$*. Consequently* MCAC *is NP-hard already in the preserving regime [S8], the greedy algorithm achieves an* $H(|Q|)$ *approximation [S9], and no polynomial-time algorithm can do asymptotically better unless* $\mathrm{P} = \mathrm{NP}$ *[S10].*

Proof: For the identity: each actor's own contribution satisfies $\hat{\sigma}_S(a) \supseteq \sigma(a) \cap \pi = \sigma(a)$ by Lemma 3.1 of the main text with $\pi = \mathcal{C}$, giving $\supseteq$; conversely, the upper bound $\hat{\sigma}_S(a) \subseteq \hat{\sigma}_S(p) \cup \sigma(a)$ and induction along each component give $\hat{\sigma}_S(a) \subseteq \bigcup_{b \in S} \sigma(b)$. The identity makes the constraint $\hat{\sigma}_S \supseteq Q$ read $\bigcup_{a \in S}(\sigma(a) \cap Q) \supseteq Q$: a weighted set cover instance verbatim. In the other direction, any weighted set cover instance embeds: take a root r with $c(r) = 0$ and $\sigma(r)$ a fresh dummy, attach one actor per set as a child of r through union edges, and let Q be the universe. The cited hardness, guarantee, and threshold results transfer directly. □

Proposition 6.6 (Target variant: polynomial on trees for all four operators (restated))**.** MCAC-T *is solvable in time* $O(d \cdot |\mathcal{C}|)$*, where* d *is the depth of the target* t*—for arbitrary mixtures of the four fundamental operators.*

Proof: By (4 of the main text), $\hat{\sigma}_S(t)$ depends only on the ancestors of t that lie in t's component of F_S: on a tree these form a contiguous chain from t upward, ancestors above a gap belong to other components, and non-ancestors never enter the recursion at t—propagation is parent-to-child, so no credential reaches t from a sibling or a descendant through any operator, union included (a union edge (p,t) contributes $\hat{\sigma}_S(p) \cup \sigma(t)$, and $\hat{\sigma}_S(p)$ depends on p's ancestors alone). Since every actor outside this chain adds cost without affecting $\hat{\sigma}_S(t)$, some optimal solution consists of t together with a contiguous prefix of its ancestors. There are only $d+1$ such prefixes; evaluate each by one downward pass and return the cheapest feasible one.

One subtlety deserves emphasis: feasibility is *not* monotone in the prefix length. Extending the chain by one more ancestor changes the component root and can destroy a required credential (an intersection ancestor filtering it away) or restore one; this is the synthesis form of the destructive intermediary, and it is why the algorithm must scan all prefixes rather than search for a threshold. □

Proposition 6.7 (Filtering regime: the coverage structure collapses (restated))**.** *Let $f(S) \coloneqq |Q \cap \hat{\sigma}_S|$ denote the coverage function of an* MCAC *instance. In the presence of a filtering edge:*

(i) f need not be monotone, and feasibility itself is not monotone: a feasible coalition can have infeasible supersets;
(ii) f need not be submodular: marginal coverage can strictly increase along a chain of coalitions;
(iii) consequently, while the decision version of MCAC *remains NP-complete (it contains the instances of Proposition 6.5, and a certificate is checkable in linear time by Proposition 6.1), the monotone-submodular structure underlying the greedy guarantee of the preserving regime is absent, and the approximability of* MCAC *with filtering edges is open.*

Proof: (i) The gadget of Theorem 4.1 applies verbatim: with $\sigma(a) = \{w\}$, $\sigma(b) = \{z, w\}$, a filtering edge (a, b) with $z \notin \pi$, and $Q = \{z\}$, the coalition $\{b\}$ is feasible while its superset $\{a, b\}$ is not.

(ii) Take a root x with $\sigma(x) = \{u\}$ for a dummy $u \notin Q$, a child c through an intersection edge with $\sigma(c) = \{z\}$, a second child d through a union edge with $\sigma(d) = \{z\}$, and $Q = \{z\}$. For $S' = \{c\}$: $f(S') = 1$, and adding x filters c's contribution to $\{u\} \cap \{z\} = \emptyset$, so $f(S' \cup \{x\}) = 0$ and the marginal is -1. For $S = \{c, d\} \supseteq S'$: $f(S) = 1$, and adding x still destroys c's contribution but d's survives the preserving edge, so $f(S \cup \{x\}) = 1$ and the marginal is 0. The marginal increased along $S' \subseteq S$, violating submodularity. (The mechanism is Theorem 4.2 at work: enlarging the coalition duplicated the victim's unique coverage, shrinking U_c and with it the joiner's destructive term.)

(iii) Immediate from (i), (ii), and Proposition 6.5; the greedy analysis of [S9] requires monotone submodular coverage, which (i) and (ii) refute, and non-monotone feasibility additionally defeats any algorithm that only ever grows a feasible partial solution. □

S2 Additional Results

S2.1 Walks, Groundedness, and Re-evaluation

Proposition S2 (Walk Characterization)**.** *A* walk *in F_S is a sequence $w = v_0 v_1 \cdots v_k$ of actors with $(v_{i-1}, v_i) \in E_S$ for all i (repetitions allowed). Its* value *is defined by $\mathrm{val}(v_0) \coloneqq \sigma(v_0)$ if $v_0 \in O_S$ and $\emptyset$ otherwise, and $\mathrm{val}(v_0 \cdots v_i) \coloneqq \otimes_{(v_{i-1}, v_i)}(\mathrm{val}(v_0 \cdots v_{i-1}),\, \sigma(v_i))$. Then, for every $a \in S$,*

$$\hat{\sigma}_S(a) = \bigcup \{ \mathrm{val}(w) : w \text{ a walk in } F_S \text{ ending at } a,\ |w| \le |S| \cdot |\mathcal{C}| \}. \tag{S5}$$

Walks starting at a non-origin are needed because $\otimes(\emptyset, Y) = Y \cap \pi(\otimes)$ need not be empty. On a DAG they are redundant and the union may be taken over source-to-a paths, which is the disjunctive reading of Definition S2. On cyclic graphs simple paths do not suffice: on the two-cycle $o \rightleftarrows b$ with both edges carrying union and $O = \{o\}$, $\hat{\sigma}(o) = \sigma(o) \cup \sigma(b)$, a value no simple path from o yields.

Proof: Write $W_k(a)$ for the union of $\mathrm{val}(w)$ over walks of length at most k ending at a. We show $W_{k-1}(a) \subseteq \Theta_S^k(\bot)(a) \subseteq W_k(a)$ for all $k \ge 1$. Lower bound: a walk of length 0 has value $[\sigma(a)]_{a \in O_S} \subseteq \Theta_S^k(\bot)(a)$; a walk $w'a$ with w' of length at most $k-2$ ending at p has value $\otimes_{(p,a)}(\mathrm{val}(w'), \sigma(a)) \subseteq \otimes_{(p,a)}(\Theta_S^{k-1}(\bot)(p), \sigma(a)) \subseteq \Theta_S^k(\bot)(a)$ by induction and monotonicity. Upper bound: for $k = 1$, $\otimes_{(p,a)}(\emptyset, \sigma(a)) = \mathrm{val}(pa)$ when $p \notin O_S$ and is contained in $\mathrm{val}(pa)$ otherwise; for the step, $\otimes_{(p,a)}(\Theta_S^k(\bot)(p), \sigma(a)) \subseteq \otimes_{(p,a)}(W_k(p), \sigma(a)) = \bigcup_w \otimes_{(p,a)}(\mathrm{val}(w), \sigma(a))$ by Lemma S1, the family being nonempty because it contains the trivial walk (p). Taking k equal to the stationarity index $K \le |S| \cdot |\mathcal{C}|$ of Proposition 3.2 gives $\hat{\sigma}_S(a) \subseteq W_K(a) \subseteq \Theta_S^{K+1}(\bot)(a) = \hat{\sigma}_S(a)$. On a DAG every non-origin v has a parent p, and the walk $p\,v \cdots$ dominates the walk $v \cdots$ by monotonicity, so walks may be restricted to start at sources; on a DAG walks are paths. □

Proposition S3 (Groundedness Characterizes the Least Solution)**.** *Call a solution E of* (S1) grounded *if every credential $c \in E(a)$ belongs to $\mathrm{val}(w)$ for some walk w ending at a—that is, if every effective credential has a finite derivation from an intrinsic credential along explicit edges. A solution is grounded if and only if it is the least solution.*

Proof: By Proposition S2, every walk value is contained in $\hat{\sigma}_S(a)$, so a grounded solution E satisfies $E \subseteq \hat{\sigma}_S$; and $\hat{\sigma}_S \subseteq E$ for every solution. Conversely, $\hat{\sigma}_S$ is grounded by (S5). □

Proposition S4 (Re-evaluation from the intrinsic state)**.** *Fix $\mathcal{G}$, its operators, σ, and origins $O \supseteq \mathrm{src}(\mathcal{G})$. If every non-source actor satisfies the incoming pass cover condition of Proposition 4.5, then the iteration $E^0 = \sigma$, $E^{k+1} = \Theta(E^k)$ is monotone increasing*

and converges to the least solution. Without the cover condition it can fail to converge: on the two-cycle $a \to b$ (intersection), $b \to a$ (union) with $O = \{a\}$ and $\sigma(b) \not\subseteq \sigma(a)$, the iteration alternates with period two.

Proof: Under the cover condition, $\Theta(E)(a) \supseteq \bigcup_p \otimes_{(p,a)}(\emptyset, \sigma(a)) = \sigma(a)$ for every non-source a (Proposition 4.5), and $\Theta(E)(a) \supseteq \sigma(a)$ for sources through the base term; hence $\sigma \subseteq \Theta(\sigma)$, so the iterates form an increasing chain by monotonicity, which converges in the finite lattice to a solution $E^* \supseteq \sigma$. The least solution contains σ (Theorem 4.3's lower bound applies verbatim under the cover condition) and every solution contains the least one; the increasing chain from σ stops at the least solution above σ, which is therefore the least solution. For the two-cycle, starting from $E^0 = \sigma$: $E^1(b) = \sigma(a) \cap \sigma(b)$, $E^1(a) = \sigma(a) \cup \sigma(b)$; $E^2(b) = E^1(a) \cap \sigma(b) = \sigma(b)$, $E^2(a) = \sigma(a) \cup (\sigma(a) \cap \sigma(b)) = \sigma(a)$; so $E^2 = E^0$ and the iteration has period two whenever $\sigma(b) \not\subseteq \sigma(a)$. □

Remark S1 (Implementations and flapping)**.** *Proposition S4 concerns one particular protocol: every actor starts by asserting its full intrinsic set and the network re-evaluates. That protocol can flap. The protocol that implements the semantics of Definition 3.11—nobody asserts anything until it is derived from an origin along an edge, which is how Datalog-based credential engines evaluate chains [S11], [S12]—is the Kleene iteration from $\perp$ of Proposition 3.2: it is monotone, terminates within $|S| \cdot |\mathcal{C}|$ rounds on any graph, cycles included, and returns the least solution. Flapping is therefore a property of an initialization choice, not of cyclic authorization as such; the open question of Section 7 of the main text is the convergence criterion for the intrinsic-state protocol beyond the pass-cover condition, not whether authorization on cycles can be computed.*

S2.2 Multi-Path Aggregation

The main text aggregates parents by union in the propagation map; the following definition, used by the results of Section S2.3, allows an arbitrary aggregator.

Definition S2 (Multi-Path Effective Credentials)**.** *When multiple paths connect actor a_0 to actor a_n, the effective credential set at a_n is determined by a* path aggregation rule Φ*:*

$$\hat{\sigma}(a_n \mid a_0) \coloneqq \Phi(\{\hat{\sigma}_p(a_n) : p \in \mathrm{Paths}(a_0, a_n)\}) \tag{S6}$$

where $\mathrm{Paths}(a_0, a_n)$ is the set of all directed paths from a_0 to a_n in $\mathcal{G}$. Simple choices for Φ include:

(i) $\Phi = \bigcup$ *(permissive: any path can establish a credential),*
(ii) $\Phi = \bigcap$ *(conservative: all paths must agree),*
(iii) $\Phi = \max_{\subseteq}$ *(maximal: take the path yielding the largest credential set).*

S2.3 Algebraic Properties

Definition S3 (Properties of Credential Operators)**.** *Let $\otimes : \mathcal{P}(\mathcal{C}) \times \mathcal{P}(\mathcal{C}) \to \mathcal{P}(\mathcal{C})$ be a credential operator.*

(i) ***Commutativity:*** *$S_1 \otimes S_2 = S_2 \otimes S_1$ for all S_1, S_2.*
(ii) ***Associativity:*** *$(S_1 \otimes S_2) \otimes S_3 = S_1 \otimes (S_2 \otimes S_3)$ for all S_1, S_2, S_3.*
(iii) ***Idempotency:*** *$S \otimes S = S$ for all S.*
(iv) ***Monotonicity:*** *If $S_1 \subseteq S_1'$ and $S_2 \subseteq S_2'$, then $S_1 \otimes S_2 \subseteq S_1' \otimes S_2'$.*
(v) ***Boundedness:*** *$S_1 \otimes S_2 \subseteq S_1 \cup S_2$ for all S_1, S_2 (the operator never creates credentials that neither actor holds).*

Proposition S5 (Non-associativity of the left join)**.** *The left join $\ltimes$ is not associative: there exist $S_1, S_2, S_3 \subseteq \mathcal{C}$ and relevance filters R, R' such that $(S_1 \ltimes_R S_2) \ltimes_{R'} S_3 \neq S_1 \ltimes_R (S_2 \ltimes_{R'} S_3)$.*

Proof: Let $\mathcal{C} = \{a, b, c, d\}$ and define:

$$S_1 = \{a, b\}, \quad S_2 = \{b, c\}, \quad S_3 = \{c, d\}$$
$$R = \{c\}, \quad R' = \{d\}$$

Left-associated:

$$S_1 \ltimes_R S_2 = S_1 \cup (S_2 \cap R) = \{a, b\} \cup (\{b, c\} \cap \{c\}) = \{a, b\} \cup \{c\} = \{a, b, c\}$$
$$(S_1 \ltimes_R S_2) \ltimes_{R'} S_3 = \{a, b, c\} \cup (S_3 \cap R') = \{a, b, c\} \cup (\{c, d\} \cap \{d\}) = \{a, b, c\} \cup \{d\} = \{a, b, c, d\}$$

Right-associated:

$$S_2 \ltimes_{R'} S_3 = S_2 \cup (S_3 \cap R') = \{b, c\} \cup (\{c, d\} \cap \{d\}) = \{b, c\} \cup \{d\} = \{b, c, d\}$$
$$S_1 \ltimes_R (S_2 \ltimes_{R'} S_3) = S_1 \cup \big((S_2 \ltimes_{R'} S_3) \cap R\big) = \{a, b\} \cup (\{b, c, d\} \cap \{c\}) = \{a, b\} \cup \{c\} = \{a, b, c\}$$

We have $\{a, b, c, d\} \neq \{a, b, c\}$, so $\ltimes$ is not associative. The credential d is present in the left-associated result but absent from the right-associated result.

An identical argument applies to the right join $\rtimes$ by symmetry. □

Remark S2 (Set operations versus path composition)**.** *It is important to distinguish two levels of structure. The properties in Proposition 3.1—commutativity, associativity, and others—are properties of* operators acting on credential sets *$S_1, S_2 \subseteq \mathcal{C}$. They describe whether $S_1 \otimes S_2 = S_2 \otimes S_1$ for a given operator $\otimes$.*

Path propagation (Definition 3.9 of the main text) is a different *operation: it composes potentially* different *operators along a sequence of edges. Even when each individual operator is commutative (e.g., $\cap$ on every edge), the path composition is* not *commutative in the actors: the credential set accumulated after visiting a_1 then a_2 may differ from that obtained by visiting a_2 then a_1, because different edges carry different operators and each actor contributes different intrinsic credentials.*

In short: commutativity of an operator $\otimes$ means the two operands *S_1 and S_2 can be swapped. It does* not *mean the* actors along a path *can be reordered. Under the tree assumption (Section S4.3), the path order is fixed by the tree structure, so this distinction is enforced by construction.*

Remark S3 (Origin of the relevance filter)**.** *The relevance filter $R \subseteq \mathcal{C}$ raises a natural question: who determines it? Several sources are possible in practice:*

(i) Exogenous (regulatory or contractual). *R is fixed by external authority—regulation, data-sharing agreements, or organizational policy. The actors cannot influence it. This is the dominant case in compliance-driven domains: for example, cross-border data-sharing agreements specify exactly which credential types may be exchanged (see Example 3, Appendix S8).*

(ii) Negotiated. *R is the outcome of a pre-game negotiation between the two actors on the edge, potentially modeled as a separate bargaining stage before the disclosure game.*

(iii) Set by the primary actor. *In a left join $S_1 \ltimes_R S_2$, the primary actor (holder of S_1) chooses R, deciding what categories of credentials it is willing to accept from the secondary actor.*

(iv) Set by the secondary actor. *The contributing actor chooses R as part of its disclosure strategy, offering only credentials it considers relevant.*

Each source implies a different game-theoretic treatment. In this paper, we treat R as exogenous and fixed per edge, deferring the strategic analysis of relevance filter determination to future work. This simplification is reasonable for the compliance-driven finance examples we develop, where regulatory and contractual constraints dominate.

Remark S4 (Two evaluation modes)**.** *The worked examples instantiate the coalition value in two distinct ways, which we now make explicit. Under* shared-outcome *evaluation, $v(S) = W(\hat{\sigma}_S)$ aggregates the effective sets of all coalition members; this is the mode used in the coalition analyses of Examples 2–4 (and in the full computation of Appendix S10). Under* at-target *evaluation, $v(S) = W(\hat{\sigma}_S(t))$ measures the effective set at a designated target actor t (the actor who must ultimately act on the assembled credentials), with $v(S) = W(\emptyset)$ when $t \notin S$; Example 1 reads its chain this way, at the last actor. The two modes can disagree sharply: an intersection edge leaves the parent's own contribution intact under shared-outcome evaluation but shrinks the value* at the child *under at-target evaluation (in Example 2, $\Delta_h(\{a_2\}) = +1$ under the former and -3 at target a_2). The structural results of Section 4.1 of the main text are stated for shared-outcome evaluation; Remark S5 records what carries over to at-target evaluation.*

S2.4 Aggregators and Evaluation Modes

Two levels should be kept apart in what follows. Results about the *algebra* concern effective credential sets: monotonicity of $\hat{\sigma}_S$ in S, and its independence of the aggregator. Results about the *game* concern the valuation W applied to those sets (Definition 3.12 of the main text). Monotonicity of $\hat{\sigma}_S$ transfers to $v(S) = W(\hat{\sigma}_S)$ for every monotone W, but superadditivity of v is a property of W, not of the algebra: Proposition S8 below requires W superadditive, and cardinality is not (two actors holding the same credential contribute it once).

Proposition S6 (Aggregator Independence of the Preserving Regime)**.** *Call an aggregator Φ* interpolating *if its value always lies between the intersection and the union of its arguments, $\bigcap_i X_i \subseteq \Phi(X_1, \ldots, X_k) \subseteq \bigcup_i X_i$, and* monotone *if it is monotone in each argument; both simple choices of Definition S2 qualify. Replace the union over parents in* (S1) *by such a Φ. If every edge with head outside O is preserving, then for every coalition S and every actor $a \in S$: $\hat{\sigma}_S(a) \supseteq \sigma(a)$ and $\hat{\sigma}_S = \bigcup_{a \in S} \sigma(a)$. Consequently, under shared-outcome evaluation the preserving regime is immune to the choice of aggregator: coalition value is monotone for every monotone interpolating Φ, not only for $\Phi = \bigcup$.*

Proof: No acyclicity is needed. With Φ monotone in each argument the propagation map Θ_S is monotone on the finite lattice $\mathcal{P}(\mathcal{C})^S$, so its least solution exists and is the limit of the increasing chain $\Theta_S^k(\bot)$ (Proposition 3.2; finiteness replaces any continuity requirement). The two bounds are proved along that chain, not along paths. Lower bound: already $\Theta_S(\bot)(a) \supseteq \sigma(a)$ for every $a \in S$—for $a \in O_S$ by the base term, and for $a \notin O_S$ because every term $\otimes_{(p,a)}(\emptyset, \sigma(a))$ contains $\sigma(a)$ (the edge is preserving), hence so does the intersection of the terms and therefore Φ of them; monotonicity of the chain preserves the inclusion in the limit. Upper bound: by induction on k, $\Theta_S^k(\bot)(a) \subseteq \bigcup_{b \in S} \sigma(b)$, since each term is contained in $X \cup \sigma(a)$ by the upper pass bound and Φ does not exceed the union of its arguments. Uniqueness of the solution, which fails on cyclic graphs, is never used: the statement concerns the least solution, and both bounds hold for it. □

Corollary S1 (The Dichotomy Is Aggregator-Independent)**.** *Under shared-outcome evaluation with any monotone interpolating aggregator, the equivalence of Theorem 4.3 holds verbatim.*

Proof: The forward direction is Proposition S6. The converse gadget of Theorem 4.3 gives b a single parent, on which every interpolating aggregator acts as the identity. □

Proposition S7 (Conjunctive Aggregation Destroys at the Target)**.** *Under* at-target *evaluation (Remark S4) with the conjunctive aggregator $\Phi = \bigcap$, monotonicity can fail even when every edge is preserving.*

Proof: Take the diamond DAG $r \to p_1 \to t$, $r \to p_2 \to t$ with all four edges carrying union, and $\sigma(r) = \{q\}$, $\sigma(p_1) = \{u\}$, $\sigma(p_2) = \{w\}$, $\sigma(t) = \{s\}$, all distinct. For $S = \{r, p_1, t\}$ there is a single source-to-t path and $\hat{\sigma}_S(t) = \{q, u, s\}$, so $v(S) = 3$. Adding p_2 creates a second path with value $\{q, w, s\}$, and conjunctive aggregation gives $\hat{\sigma}_T(t) = \{q, u, s\} \cap \{q, w, s\} = \{q, s\}$, so $v(T) = 2 < 3$. The destruction resides entirely in the aggregation rule: no edge filters anything. □

Remark S5 (Beyond trees, leaves, and shared outcomes)**.** *Three boundaries of the results above deserve note.* General graphs. *Section 4.3 of the main text lifts the coalition semantics and the full dichotomy—both directions—to arbitrary directed graphs, cycles included, under a least-solution semantics that reduces to the permissive aggregator $\Phi = \bigcup$ of Definition S2 on acyclic graphs. Under the conservative aggregator $\Phi = \bigcap$ the picture splits by evaluation mode: shared-outcome value in the preserving regime is immune to the aggregator altogether (Proposition S6), but* at-target *value can fail monotonicity even when every edge is preserving (Proposition S7): adding an actor adds paths, and requiring all paths to agree shrinks the set at the target. This conjunctive-veto pathology has a long pedigree in the economics of the anticommons [S13]; unlike the anticommons, whose vetoes gate* consent *to use a resource, the pathology here operates on credential* content*, and it resides in the aggregation rule rather than on any edge.* Non-isolated filtered children. *When a filtered child carries a subtree, the shrunken input propagates downward and equation* (S3) *relaxes to the two-sided bound $\hat{\sigma}_T \setminus \hat{\sigma}_S \subseteq \hat{\sigma}_T(x) \setminus \hat{\sigma}_S$ and $\hat{\sigma}_S \setminus \hat{\sigma}_T \subseteq \bigcup_i U_i'$ over subtree unique coverage; exactness fails because a credential lost on one branch may be resupplied along another.* At-target evaluation. *Under at-target evaluation (Remark S4) both Theorem 4.1 and Proposition 4.2 survive: a joiner with no children in S cannot lie on the path to the target unless it is the target itself, and with every edge preserving, placing x above the target's component root only enlarges every set along the path (lower pass bound and monotonicity). What changes is* who *destroys and* where*: an interior joiner above the target through a filtering edge shrinks the target's own set—in Example 2 with target a_2, $\Delta_h(\{a_2\}) = |\{c_2\}| - |\sigma(a_2)| = -3$—and the destroyed credentials are read off the single path from the component root to the target, so the accounting is exact but concerns one actor's set rather than the coalition-wide union.*

Remark S6 (Why the least solution)**.** *On trees and DAGs the solution is unique, so nothing is chosen. On cyclic graphs the equations can have many solutions: for the two-cycle $a \to b$ with intersection and $b \to a$ with union, $O = \{a\}$, every assignment $E(a) = \sigma(a) \cup T$, $E(b) = E(a) \cap \sigma(b)$ with $T \subseteq \sigma(b) \setminus \sigma(a)$ is a solution. Three independent considerations single out the least one.* Operational. *Definition 3.11 of the main text is not an equation but a process: credentials flow from the origin along edges, and an actor's own credentials enter only through its incoming edge. The least solution is the closure of that process on an arbitrary graph—the limit of propagation from the state in which origins hold their intrinsic sets and nobody else holds anything—and it is the only solution reachable from that state. The seemingly natural alternative, starting from every actor holding its intrinsic set and iterating Θ_S, defines nothing: on the two-cycle above it oscillates with period two.* Groundedness. *Every non-least solution contains credentials that are supported only circularly—T is in $E(a)$ because it is in $E(b)$, and in $E(b)$ because it is in $E(a)$—which in authorization is authority conjured from a cycle of mutual assumption; Proposition S3 says the least solution is exactly the one free of such credentials. This is the same reason for which Datalog-based trust-management languages adopt least-model semantics for credential chains [S11], [S12].* Conservativeness. *The least solution is pointwise below every other, so $v(S)$ under it is a lower bound on $v(S)$ under any consistent semantics: coalitions are credited only with what they can derive, a default-deny reading; the greatest solution, by contrast, degenerates (on a union two-cycle without origins it equals $\mathcal{C}$ for every σ). Note that minimality concerns levels, not differences: marginal contributions and Shapley values are differences of values, and no ordering between solutions carries over to them.*

Remark S7 (Origins, cycles, and the rooting assumption)**.** *Proposition 4.5 explains the standing assumption that the actor graph is rooted at the query origin (Section S4.3). On a rooted tree the root is the unique source, so the origin is structurally forced and nothing is assumed; on a graph in which a strongly connected component of two or more actors has no entering edge, no structural source exists there, and a designated origin must be supplied—but it changes the outcome only in the filtering regime. Note that the cover condition is weaker than "every edge preserving": an actor with one union parent and one intersection parent is covered. The coalition-quantified dichotomy of Theorem 4.3 nevertheless requires every edge, because a coalition may drop the union parent. Two quantifiers, two exact criteria, and the same fault line.*

Remark S8 (What does not lift)**.** *Three boundaries remain. The* exact *criterion of Theorem 4.2 degrades beyond trees to the two-sided bounds of Remark S5: a credential lost through one filtered parent may be resupplied along another walk, and the accounting of unique coverage no longer closes. Aggregation is not a boundary for shared-outcome value (Corollary S1), but the conjunctive pathology of Proposition S7 lives in at-target evaluation. And target-directed synthesis (*MCAC-T*, Section 6 of the main text) relies on the unique ancestor chain of a tree. Cycles, by contrast, are no longer a boundary for semantics, evaluation, the operator-level dichotomy, harmlessness of sinks, or hardness; they are a boundary for the assignment-level audit, which is linear up to acyclic graphs and coNP-complete beyond (Propositions S1 and 4.3). What the least-solution semantics does not settle is the* dynamics *of repeated re-evaluation on cycles, which we return to among the open problems (Section 7 of the main text). Within these boundaries the division of labor is clean: semantics, evaluation, the dichotomy, and all hardness results live on arbitrary directed graphs; the protection audit lives on acyclic graphs; exact destructiveness accounting and target-directed synthesis are where trees still earn their keep.*

S2.5 Credential Protection: Placement and Cycles

The tree-level protection criterion (Proposition 4.1 of the main text) and its general-graph form (Proposition S1) are proved in Section S1; the two remarks below record their design reading and the two independent sources of hardness.

Remark S9 (Protection as a placement rule)**.** *Theorem 4.1 is Proposition 4.1(ii) quantified over assignments: if (p, h) is filtering with $z \notin \pi$, the gadget assignment $\sigma(h) = \{z, w\}$, $\sigma(p) = \{w\}$ of the theorem makes h an unprotected holder of z, and if every edge is preserving, every holder is protected through the clause $c \in \pi = \mathcal{C}$. The proposition thus separates two things the dichotomy bundles: the operators, which say whether* some *assignment can be harmed, and the placement of credentials, which says whether* this *one can. Its practical reading is a rule for redundancy. A second copy of a credential guards it in every coalition that contains the copy* if the copy is protected*; an unprotected copy guards nothing and adds a target, since by (ii) it creates a coalition in which the credential is lost. The same proposition draws a line inside the complexity results of Section 6 of the main text: certifying that a deployment can never lose a credential is linear, whereas certifying that a given actor never* nets *a loss against its gains is coNP-complete (Theorem 6.1); the hardness lies in the netting, not in finding the exposure.*

Remark S10 (Two sources of hardness)**.** *Theorem 6.1 obtains hardness on trees of depth one from many credentials: with a single actor's gains netted against the losses it causes, the difficulty is combinatorial in $\mathcal{C}$, and Proposition 6.3 confirms that $|\mathcal{C}|$ is the whole source of it on trees. Proposition 4.3 obtains hardness from cycles alone, with one credential and no netting: the difficulty is that a coalition can starve a credential only by closing a loop of filters around all of its holders, and choosing the loop consistently is a satisfiability problem. On acyclic graphs neither source is present for the assignment-level question (Proposition S1(ii)); whether* DESTRUCTIVE-ACTOR *remains fixed-parameter tractable in $|\mathcal{C}|$ on acyclic graphs is open.*

S2.6 Disclosure Equilibria

Lemma S3 (Locality of the fundamental operators)**.** *Each of the four fundamental operators is* local*: for every credential $c \in \mathcal{C}$, membership of c in $\otimes(S_1, S_2)$ is determined by the pair $(\mathbf{1}[c \in S_1], \mathbf{1}[c \in S_2])$ alone (together with the fixed constant $\mathbf{1}[c \in R]$ for the joins), via a Boolean function that is monotone in both arguments. Explicitly, $c \in S_1 \cap S_2$ if and only if $\mathbf{1}[c \in S_1] \wedge \mathbf{1}[c \in S_2]$; $c \in S_1 \cup S_2$ if and only if $\mathbf{1}[c \in S_1] \vee \mathbf{1}[c \in S_2]$; $c \in S_1 \ltimes_R S_2$ if and only if $\mathbf{1}[c \in S_1] \vee (\mathbf{1}[c \in S_2] \wedge \mathbf{1}[c \in R])$; and symmetrically for $\rtimes_R$.*

Proof: Immediate from the defining formulas in Definition 3.6 of the main text; monotonicity holds because each formula is a composition of conjunctions and disjunctions without negation. □

Corollary S2 (Join edges)**.** *Let the edge operator be $\ltimes_R$ (symmetrically $\rtimes_R$), let $\sigma(a) \cap (\sigma(b) \cap R) = \emptyset$, let each V_i be supermodular, and let each D_i be submodular* and nondecreasing *with respect to set inclusion. Then $\Gamma_{a,b}$ admits a pure-strategy Nash equilibrium.*

Proof: By Lemma S3, credentials in $\sigma(b) \setminus R$ never enter the effective set, so for b, dropping them from any disclosed set leaves V_b's argument unchanged and, by monotonicity of D_b, weakly lowers cost. Consider the *reduced game* in which b's strategy space is $\mathcal{P}(\sigma(b) \cap R)$; on this domain $\delta_a \ltimes_R \delta_b = \delta_a \cup \delta_b$, so the reduced game is a union game satisfying the hypotheses of Theorem 5.1 (with ground sets $\sigma(a)$ and $\sigma(b) \cap R$), and admits a pure equilibrium (δ_a^*, δ_b^*). We claim this profile is an equilibrium of the full game. Actor a's strategy space and payoffs are unaffected by the reduction, so a has no profitable deviation. For b, any deviation $\delta_b' \subseteq \sigma(b)$ satisfies $u_b(\delta_a^*, \delta_b') = V_b(\delta_a^* \cup (\delta_b' \cap R)) - D_b(\delta_b') \leq V_b(\delta_a^* \cup (\delta_b' \cap R)) - D_b(\delta_b' \cap R) \leq u_b(\delta_a^*, \delta_b^*)$, where the first inequality is monotonicity of D_b and the second is the equilibrium property in the reduced game. □

Remark S11 (Existence in mixed strategies)**.** *Since the strategy sets S_a and S_b are finite (power sets of finite credential sets), the disclosure game is a finite game and admits at least one Nash equilibrium in mixed strategies by Nash's theorem [S14]. This guarantee is unremarkable in itself; it matters here only because* pure*-strategy equilibria can genuinely fail to exist in disclosure games, as we now show.*

Remark S12 (Equilibrium structure by operator)**.** *Proposition 5.1 exhibits the best-response cycle characteristic of matching-pennies-type games, arising here from the combination of credential complementarity on one side of the edge with credential substitutability on the other. In benign special cases pure equilibria do exist: when the operator is intersection and disclosure is costless, full disclosure is an equilibrium (withholding only shrinks the intersection); under union with high disclosure costs, prisoners'-dilemma-like dynamics arise in which mutual withholding is a pure but socially suboptimal equilibrium. In the limiting case $D_i \to \infty$ for all actors—a formal expression of a never-disclose-without-necessity policy—equilibrium collapses to the minimal overlap: each actor contributes only what the operator and the other party's credentials force into the output. We explore specific equilibrium structures in the examples; a systematic classification of when pure equilibria are guaranteed to exist follows in Section 5.2 of the main text.*

Remark S13 (The classification, tightness, and an open case)**.** *The results of this subsection combine into the following picture ("exists" means a pure-strategy Nash equilibrium is guaranteed; "may fail" means a counterexample exists):*

Preference structure	$\cap$	$\cup$	$\ltimes_R, \rtimes_R$
Arbitrary V_i	*exists (Prop. 5.2)*	*may fail (Prop. 5.1)*	*may fail*
Modular V_i, D_i	*exists*	*exists*	*exists (Prop. 5.3)*
Supermodular V_i (disjoint holdings)	*exists*	*exists (Thm. 5.1)*	*exists (Cor. S2)*
Common interest ($V_a = V_b$)	*exists*	*exists*	*exists (Prop. 5.4)*

The counterexample of Proposition 5.1 is tight for this classification in three respects. First, it transfers verbatim to the join operators: setting $R = \mathcal{C}$ makes $\ltimes_R$ coincide with union, which justifies the "may fail" entries in the last column. Second, its value functions are strictly *of opposite modularity type—V_a is supermodular but not modular (bundle-only value), V_b is submodular but not*

modular (first-credential-only value)—so it shows that neither the both-supermodular hypothesis of Theorem 5.1 nor the modularity hypothesis of Proposition 5.3 can be dropped. Third, it has $V_a \neq V_b$, showing the common-interest hypothesis of Proposition 5.4 cannot be dropped either. Non-existence thus requires simultaneously: an expansive operator, non-modular valuations, heterogeneous interests, and opposite complementarity structure on the two sides of the edge.

One region of the map remains open: union edges where both *value functions are submodular (both actors view credentials as substitutes) with multi-credential holdings. When each actor holds a single credential, the resulting 2×2 game is of the chicken type and a pure equilibrium exists, but for larger holdings the lattice-theoretic machinery of Theorem 5.1 does not apply (submodular own-utility breaks supermodularity in one's own strategy), and we know of neither a general existence proof nor a counterexample. We record this as an open problem.*

S2.7 Coalition Formation

Remark S14 (Edge-level equilibria versus the coalition optimum)**.** *The existence results of Section 5.2 of the main text are statements about the bilateral game on one edge, and their proofs use locality (Lemma S3): on a single edge the fate of each credential depends only on the two disclosure decisions concerning it. Locality does not extend to the coalition: an actor's disclosure changes the context of every actor below it, so at depth two or more the fate of a credential depends on decisions along the whole path. The characteristic function (9 of the main text) therefore does not decompose into edge games, and no equilibrium claim is made for it. Its maximizing disclosure profile δ^* exists simply because the strategy space is finite; it is a cooperative optimum, not a Nash equilibrium of any non-cooperative disclosure game on the graph, and the coalition game is analyzed through v alone.*

Proposition S8 (Superadditivity under Union)**.** *If all edge operators are $\cup$ (union), disclosure costs are zero, and the global value function W is* superadditive*—that is, $W(A \cup B) \geq W(A) + W(B)$ for all disjoint $A, B \subseteq \mathcal{C}$, with $W(\emptyset) = 0$—then:*

(i) Monotonicity: $v(S \cup T) \geq \max\big(v(S),\, v(T)\big)$ *for all disjoint $S, T \subseteq \mathcal{A}$.*

(ii) Superadditivity: $v(S \cup T) \geq v(S) + v(T) - W(\hat{\sigma}_S \cap \hat{\sigma}_T)$ *for all disjoint $S, T \subseteq \mathcal{A}$. In particular, if the effective credential sets $\hat{\sigma}_S$ and $\hat{\sigma}_T$ are disjoint, then $v(S \cup T) \geq v(S) + v(T)$.*

Proof: Under union with zero cost, $\hat{\sigma}_{S\cup T} \supseteq \hat{\sigma}_S \cup \hat{\sigma}_T$. Since W is superadditive with $W(\emptyset) = 0$, it is monotone: for $A \subseteq B$, writing $B = A \cup (B \setminus A)$ gives $W(B) \geq W(A) + W(B \setminus A) \geq W(A)$.

Part (i). Monotonicity of W and $\hat{\sigma}_{S\cup T} \supseteq \hat{\sigma}_S$ give $v(S \cup T) = W(\hat{\sigma}_{S\cup T}) \geq W(\hat{\sigma}_S) = v(S)$, and similarly for T.

Part (ii). Decompose $\hat{\sigma}_S \cup \hat{\sigma}_T = \hat{\sigma}_S \cup (\hat{\sigma}_T \setminus \hat{\sigma}_S)$ into disjoint parts. By superadditivity of W:

$$W(\hat{\sigma}_S \cup \hat{\sigma}_T) \geq W(\hat{\sigma}_S) + W(\hat{\sigma}_T \setminus \hat{\sigma}_S)$$

By monotonicity, $W(\hat{\sigma}_T) \leq W(\hat{\sigma}_T \setminus \hat{\sigma}_S) + W(\hat{\sigma}_S \cap \hat{\sigma}_T)$, so $W(\hat{\sigma}_T \setminus \hat{\sigma}_S) \geq W(\hat{\sigma}_T) - W(\hat{\sigma}_S \cap \hat{\sigma}_T)$. Combining with monotonicity of W on $\hat{\sigma}_{S\cup T} \supseteq \hat{\sigma}_S \cup \hat{\sigma}_T$:

$$v(S \cup T) \geq W(\hat{\sigma}_S) + W(\hat{\sigma}_T) - W(\hat{\sigma}_S \cap \hat{\sigma}_T) = v(S) + v(T) - W(\hat{\sigma}_S \cap \hat{\sigma}_T)$$

When $\hat{\sigma}_S \cap \hat{\sigma}_T = \emptyset$, we get $v(S \cup T) \geq v(S) + v(T)$. □

Remark S15 (Disconnected coalitions)**.** *When the induced subgraph on S is disconnected—i.e., some actors in S share no edges—each connected component computes its effective credentials independently: no credential operator combines credentials across components, since no edge exists to carry one, and the roots of the components contribute their intrinsic sets (Definition 3.11 of the main text). Under shared-outcome evaluation the coalition effective set is the union over components, so $v(\{a_i, a_j\}) = |\sigma(a_i) \cup \sigma(a_j)|$ for two isolated actors; this equals $|\sigma(a_i)| + |\sigma(a_j)|$ exactly when the two intrinsic sets are disjoint (as in Example 2) and is smaller when they overlap. Under at-target evaluation only the component of the target matters.*

Remark S16 (Practical use of the bound)**.** *Part (ii) of Proposition S8 requires knowledge of $\hat{\sigma}_S \cap \hat{\sigma}_T$, the credential overlap between coalitions. In practice, this overlap is often estimable: system designers know which actors hold similar certifications (e.g., two agents both holding anti-money-laundering (AML) clearance), or can bound it from the intrinsic credential sets since $\hat{\sigma}_S \subseteq \bigcup_{a\in S} \sigma(a)$. The bound is most useful in two cases: (a) when actors are specialized (small overlap, bound is tight), and (b) as a diagnostic tool for identifying which credential overlaps are most costly to the coalition structure.*

S2.8 Synthesis and the Complexity Landscape

Remark S17 (Synthesis and the dichotomy)**.** *The three results align synthesis with the fault line of Theorem 4.1. In the preserving regime the algebra is invisible—$\hat{\sigma}_S$ collapses to a union of intrinsic sets—and synthesis is* classically *hard: NP-hard but monotone, submodular, and greedily approximable to the optimal threshold. The target-directed question an AI-workflow orchestrator asks ("deliver requirement Q to actor t") is outright polynomial on trees, for any operator mixture. It is the combination of shared-outcome coverage with filtering edges that is* wildly *hard: there the approximation scaffolding collapses, and even feasibility stops being monotone. On general graphs, the target variant acquires a Steiner-tree flavor under multi-path aggregation; we leave its complexity, and the approximability question of part (iii), open.*

Remark S18 (The complexity landscape)**.** *The three results delimit what is tractable. Evaluating any single coalition—and hence auditing any* proposed *configuration—is fast (linear); this is what the* 63*-coalition computation of Appendix S10 exercises exhaustively at small scale. What is hard is* global structural risk detection*: certifying that an actor is never destructive across all*

possible coalitions is coNP-complete by Theorem 6.1, and the hardness arises already from the interaction of a single intersection edge with union edges—the minimal mixed-operator setting. Two weaker certificates remain cheap: certifying that a deployment can never lose *a credential—exposure without netting—is linear on trees and acyclic graphs (Propositions 4.1 and S1), and the exact decision is fixed-parameter tractable in $|\mathcal{C}|$ on trees (Proposition 6.3); both cheap routes end at cycles, where even the single-credential question is coNP-complete (Proposition 4.3). Value attribution occupies a middle ground: exact Shapley computation inherits the exponential coalition space, but—unlike cooperative games in which even the value oracle is hard [S6]—the polynomial evaluation on trees makes Monte Carlo estimation practical at any scale, with explicit sample bounds. Practically: audit configurations exactly, certify safety only for restricted operator regimes (Theorem 4.1), attribute value by sampling—and synthesize coalitions greedily in the preserving regime, by prefix scan for target-directed authorization, and with no known guarantees once filtering edges enter (Propositions 6.5–6.7).*

S3 Mechanism Design for Credential Elicitation

In many practical settings, a *principal* (e.g., an orchestration platform or workflow manager) must elicit truthful credential disclosure from strategic actors who may misreport their credentials to reduce disclosure costs or gain a strategic advantage. The challenge is to design rules—an allocation of effective credentials and a schedule of transfers—that make honest reporting each actor's best response.

Definition S4 (Credential Type)**.** *Each actor a has a private* type $\theta_a = \sigma(a)$*, which is the actor's true intrinsic credential set. The type is private information: other actors and the principal observe only what a chooses to disclose.*

Definition S5 (Credential Mechanism)**.** *A* credential mechanism *is a tuple $M = (f, t)$ where:*

(i) $f : \prod_{a \in \mathcal{A}} \mathcal{P}(\mathcal{C}) \to \mathcal{P}(\mathcal{C})$ *is the* allocation rule *that takes the reported credential sets from all actors and produces the effective credential set for the workflow.*

(ii) $t : \prod_{a \in \mathcal{A}} \mathcal{P}(\mathcal{C}) \to \mathbb{R}^{|\mathcal{A}|}$ *is the* transfer rule *that assigns a payment (positive or negative) to each actor based on all reports.*

Definition S6 (Incentive Compatibility)**.** *A mechanism $M = (f, t)$ is* incentive compatible *(IC) if for every actor $a \in \mathcal{A}$ and every pair of types θ_a, θ'_a:*

$$V_a(f(\theta_a, \theta_{-a})) + t_a(\theta_a, \theta_{-a}) - D_a(\theta_a) \geq V_a(f(\theta'_a, \theta_{-a})) + t_a(\theta'_a, \theta_{-a}) - D_a(\theta'_a) \tag{S7}$$

for all θ_{-a}, where θ_{-a} denotes the types of all actors other than a. That is, truthful disclosure maximizes each actor's utility.

Definition S7 (Individual Rationality)**.** *A mechanism is* individually rational *(IR) if every actor's utility from participation is non-negative under truthful reporting:*

$$V_a(f(\theta_a, \theta_{-a})) + t_a(\theta_a, \theta_{-a}) - D_a(\theta_a) \geq 0 \quad \forall a \in \mathcal{A} \tag{S8}$$

Proposition S9 (VCG Mechanism for Credential Elicitation)**.** *If the allocation rule f maximizes total welfare $\sum_{a \in \mathcal{A}} V_a(f(\boldsymbol{\theta}))$ and the transfer rule is:*

$$t_a(\boldsymbol{\theta}) = \sum_{j \neq a} V_j(f(\boldsymbol{\theta})) - \sum_{j \neq a} V_j\big(f^{-a}(\boldsymbol{\theta}_{-a})\big) \tag{S9}$$

where f^{-a} is the welfare-maximizing allocation without actor a, then the mechanism is incentive compatible and individually rational.

Proof. This is the standard VCG result applied to our setting. The transfer t_a equals the externality that a's participation imposes on all other actors. Under this transfer, each actor's dominant strategy is truthful revelation of $\theta_a = \sigma(a)$. Individual rationality follows from the non-negativity of marginal contributions when W is monotone. See [S15] for the general proof. □

Remark S19. *The VCG mechanism provides a theoretical benchmark but has known practical limitations. It requires a* trusted principal*—an entity outside the strategic game that can observe reported credentials, compute the welfare-maximizing allocation, and enforce transfers. In practice, this role could be filled by an orchestration platform, a regulatory body, or a cryptographic protocol (as in the DAO-Agent framework [S16], which uses zero-knowledge proofs to verify coalition values without a centralized authority). The principal is distinct from the actors in the graph: it does not hold credentials or participate in propagation, but sets the rules under which actors disclose. Further limitations include budget imbalance and computational intractability for large actor graphs. We explore more practical mechanisms in Example 3 (Appendix S8), where regulatory constraints further restrict the mechanism design space.*

Remark S20 (Limitations of the VCG layer)**.** *Two classical caveats apply and should temper any literal reading of the mechanism. First,* budget imbalance*: efficient dominant-strategy mechanisms generically cannot balance their transfers [S17], so an external orchestrator must inject or absorb payments. In intra-organizational agent networks, monetary transfers to internal components are unnatural; the transfers of Proposition S9 are therefore best read as an accounting or priority currency, or the mechanism used as an* audit benchmark*—the transfer an actor* would *owe quantifies its strategic externality—rather than as a literal payment scheme. Second, a* privacy tension*: truthful revelation asks each actor to disclose its full credential set to the mechanism, in direct tension with the minimal-disclosure doctrine whose equilibrium logic Proposition 5.2 formalizes at the edge level. Cryptographic verification of coalition values (as in the zero-knowledge approach of [S16]) mitigates what the* center *must learn; designing disclosure-minimal elicitation mechanisms for credential games remains open. Two further limitations are inherited from VCG itself and are not specific to*

credentials: the mechanism is robust to unilateral deviations only, so colluding actors can misreport jointly, and an actor that can split itself into several virtual agents (a Sybil attack) can inflate its apparent marginal contribution and its transfer. In cross-organizational settings, where one organization controls several agents, both are live threats—a further reason to read the mechanism as an audit benchmark rather than as a deployable protocol.

S4 Simplifying Assumptions

The general framework of Section 3 of the main text admits arbitrary credential spaces, graph topologies, and operator assignments; the theory of Sections 4 of the main text–6 of the main text was developed under standing assumptions stated where they are used, and the worked examples operate under the same ones. This section collects those seven simplifying assumptions in one place and defends them. For each, we state the assumption, discuss what generality is lost, note where the present paper already relaxes it, and indicate how it could be relaxed further.

S4.1 Binary, Atomic Credentials

Assumption. Credentials are binary and atomic: for each actor a and credential c, either $c \in \sigma(a)$ or $c \notin \sigma(a)$, and credentials are unstructured identifiers with no implication between them.

The general framework permits a richer formulation in which credentials carry *degrees*, with $\sigma(a, c) \in [0, 1]$ representing the strength or confidence of an attestation. The fundamental operators then generalize via the standard Zadeh extension [S18] of set operations to fuzzy sets (intersection becomes pointwise min, union becomes pointwise max). We adopt the binary simplification because it keeps the examples combinatorially transparent; the fuzzy generalization is a natural direction for future work.

Atomicity deserves a separate word, because deployed agent permissions are often *parameterized*—"may transfer if the amount is below $1,000 and the currency is USD"—and one may fear that decomposing such predicates into atoms blows up $|\mathcal{C}|$ and with it the fixed-parameter algorithms of Section 6 of the main text. Two observations bound the concern. First, an atom is an *attestation*, not an instantiation: the parameterized permission above is one credential, held or not, exactly as a signed capability token is one token; the algebra never enumerates parameter values, and the credential universe of a workflow is the set of distinct attestations that occur in it, which in the worked examples is at most twelve and in deployed workflows is typically tens, not thousands. Second, the audit results do not depend on $|\mathcal{C}|$ being small at all—evaluation and marginal contributions are linear in $|\mathcal{C}|$ (Proposition 6.1), and so is the protection audit (Proposition 4.1)—so only exact global certification (Proposition 6.3) is sensitive to the size of the universe, as its NP-completeness says it must be. What atomicity genuinely excludes is *implication between credentials*: "below $1,000" entails "below $2,000," and a universe with such entailments is a lattice rather than a powerset. Extending the algebra from $\mathcal{P}(\mathcal{C})$ to a distributive lattice of structured permissions—where the pass filter becomes a lattice element and the protection criterion a statement about lattice ideals—is the natural next step, and we list it among the open problems.

S4.2 Actor Scores as Separate Structure

Assumption. Each actor $a \in \mathcal{A}$ carries, in addition to its credential set $\sigma(a)$, a *score vector* $s(a) \in \mathbb{R}^k$ representing reputation, reliability, and other quantitative attributes. Scores are not credentials: they do not participate in set operations. Instead, they enter the game-theoretic layer as parameters.

Definition S8 (Score Function)**.** *A* score function *is a mapping $s : \mathcal{A} \to \mathbb{R}^k$ assigning to each actor a k-dimensional score vector. Components of $s(a)$ may include reputation scores, transaction success rates, compliance ratings, uptime metrics, or any other quantitative measure of trustworthiness.*

Scores modulate the game-theoretic layer in three ways:

(i) **Value modulation.** The value that actor i derives from collaborating with actor j depends on j's score: $V_i(\hat{\sigma}, s(j))$ replaces $V_i(\hat{\sigma})$ in equation (8 of the main text). An actor with higher reputation makes the collaboration more valuable.

(ii) **Cost modulation.** An actor's disclosure cost may depend on its own score: a well-established actor (high reputation) may face lower disclosure cost because it has less to lose from transparency. Formally, $D_i(\delta, s(i))$ replaces $D_i(\delta)$.

(iii) **Operator selection.** The operator assigned to an edge (a, b) may depend on the scores of the endpoint actors. For instance, an orchestration platform might assign union to edges where both actors have scores above a trust threshold, and intersection otherwise:

$$\otimes_{a,b} = \begin{cases} \cup & \text{if } s(a)_1 \geq \tau \text{ and } s(b)_1 \geq \tau \\ \cap & \text{otherwise} \end{cases} \tag{S10}$$

where $s(\cdot)_1$ denotes the first component (e.g., reputation) and τ is a threshold.

Remark S21. *The separation of scores from credentials is a modeling choice, not a mathematical necessity. One could embed scores into the credential space by introducing "graded credentials"—effectively the fuzzy set approach of Section S4.1. We keep them separate because it mirrors the architecture of emerging agent identity systems (e.g., ERC-8004, which maintains distinct identity, reputation, and validation registries) and because it preserves the clean combinatorial structure of the credential algebra.*

S4.3 Tree Graphs Rooted at Query Origin

Assumption. The actor graph $\mathcal{G} = (\mathcal{A}, E)$ is a *directed rooted tree*: there exists a distinguished actor $a_0 \in \mathcal{A}$ (the *root*) such that for every other actor $a \in \mathcal{A}$, there is exactly one directed path from a_0 to a.

This assumption eliminates three sources of complexity in the general framework:

(i) **Cycles.** A tree has no cycles, so the convergence question for propagation (Remark after Definition 3.4 of the main text) does not arise; Section 4.3 of the main text settles it in general by the least-solution semantics.

(ii) **Multi-path ambiguity.** Each actor is reachable by exactly one path from the root, so the path aggregation rule Φ (equation (S6)) is unnecessary.

(iii) **Propagation direction.** Credentials propagate unambiguously from root to leaves, giving a natural sequential structure.

The tree assumption is practically realistic for the majority of current agentic AI workflows, which are initiated by a query and fan out through orchestration layers to specialized agents and tools. It is also the assumption this paper itself relaxes furthest: Section 4.3 of the main text lifts the coalition semantics, the evaluation algorithm, and the full monotonicity dichotomy to arbitrary directed graphs, cycles included, and the scope table there records exactly which results remain tree-bound (the exact destructiveness criterion and target-directed synthesis).

S4.4 Root Type: Human versus Agent

The tree must have a root. We consider both cases explicitly.

Case 1: Human root ($a_0 \in \mathcal{A}_H$): This is the classical scenario: a human user initiates a query, which propagates through an orchestrator to downstream agents and tools. The classical delegation model (Proposition 3.3) is a subcase where all edge operators are intersection. Our framework generalizes this by permitting non-intersection operators on any edge, allowing downstream actors to contribute credentials that the human does not hold.

Case 2: Agent root ($a_0 \in \mathcal{A}_M$): This scenario arises when an autonomous agent initiates a workflow without direct human involvement—for example, a monitoring agent detecting an anomaly and triggering a remediation chain, or a scheduling agent autonomously coordinating a multi-party process.

In the classical delegation model, this case is *undefined*: there is no human principal from whom to derive authority. In our framework, it is handled naturally. The agent root a_0 carries its own intrinsic credential set $\sigma(a_0)$, and propagation proceeds identically to the human-root case. The effective credentials at any node depend on the credentials of *all* actors along the path, not on the type of the root.

S4.5 Operators from the Four Fundamentals

Assumption. Each edge operator $\otimes_{a,b}$ is one of the four fundamental operators defined in Definition 3.6 of the main text: intersection ($\cap$), union ($\cup$), left join ($\ltimes$), or right join ($\rtimes$).

This restricts the operator assignment to a finite, well-characterized set. The structural and complexity results hold for the whole class $\mathcal{P}$ of pass-filtered operators (Definition 3.8); the four are its canonical instances, and the strategic results additionally use their locality. The properties established in Proposition 3.1 apply directly, and the behavior of each operator is intuitively interpretable in authorization terms:

Operator	When to use
$\cap$	Conservative: both actors must agree (e.g., compliance-critical steps)
$\cup$	Expansive: pool all capabilities (e.g., peer collaboration)
$\ltimes$	Asymmetric: primary actor's credentials plus filtered contribution from secondary (e.g., human delegates to agent, keeping human credentials primary)
$\rtimes$	Reverse asymmetric (e.g., agent delegates upward to human for approval)

S4.6 Sequential Disclosure

Assumption. On a rooted tree, disclosure proceeds *sequentially* from root to leaves. The root discloses first; each child observes its parent's disclosure before choosing its own.

This replaces the simultaneous disclosure game of Section 5.1 of the main text with a sequential (extensive-form) game. The solution concept changes accordingly: we seek *subgame perfect equilibria* (SPE) rather than Nash equilibria. Existence here does not rely on mixing: a finite extensive-form game of perfect information admits a pure-strategy SPE by backward induction. Notably, sequentiality thus *restores* the pure-strategy existence that can fail in the simultaneous game (Proposition 5.1); SPE further refines Nash equilibrium by eliminating equilibria that rely on non-credible threats.

Remark S22. *Sequential disclosure on a tree has several advantages over the simultaneous formulation:*

(i) It is more realistic: in practice, the initiator of a workflow reveals its requirements before downstream agents respond.

(ii) SPE eliminates non-credible threats that can arise in simultaneous Nash equilibria.

(iii) On a finite tree, the SPE can be computed by backward induction *from the leaves, which is computationally tractable.*

S4.7 Credential Complementarity

Assumption. Where stated—and only there—the global value function $W : \mathcal{P}(\mathcal{C}) \to \mathbb{R}_{\geq 0}$ is *superadditive*: for all disjoint $A, B \subseteq \mathcal{C}$,

$$W(A \cup B) \geq W(A) + W(B) \tag{S11}$$

with $W(\emptyset) = 0$.

We call this property *credential complementarity*: the value of holding two credential sets together is at least as great as the sum of their individual values. Unlike A1–A6, this assumption is *not* adopted by the worked examples: every number in Examples 1–4 is computed with $W = |\cdot|$, cardinality, which is monotone but subadditive on overlapping holdings and therefore not superadditive (Example 2: $v(\{h, a_1\}) = 6 < 5 + 3$). A7 is invoked by exactly one result, Proposition S8 (superadditivity of the coalition game under union), and by the illustrative valuation W_{comp} of Example 1; all structural, complexity, and monotonicity results require only that W be monotone.

The structure of W reflects the nature of the credential system. Three regimes are possible:

(i) **Complementary (W superadditive).** Credentials unlock value in combination. A credential that is worthless in isolation may be essential when combined with others. This is the natural regime for authorization systems where access control is *conjunctive*—an action requires multiple credentials simultaneously. In the finance domain: KYC verification (c_1) alone does not permit trading, and trading authorization (c_2) alone does not satisfy regulatory requirements, but together they enable a compliant trade. The value of the pair exceeds the sum of the parts.

(ii) **Independent (W additive).** Each credential contributes a fixed value regardless of what other credentials are present: $W(A \cup B) = W(A) + W(B)$ for disjoint A, B. This arises when credentials enable independent capabilities with no interaction. Example: holding a license to trade equities and a separate license to trade commodities, where the two markets are unrelated.

(iii) **Redundant (W subadditive).** Overlapping credentials provide diminishing returns: $W(A \cup B) \leq W(A) + W(B)$. This arises when multiple credentials certify similar capabilities. Example: two AML certifications from different jurisdictions—the second adds some coverage but less than its standalone value, since much of the compliance ground is already covered.

Remark S23. *In practice, real credential systems exhibit a mix of all three regimes across different credential pairs. Complementarity is the regime in which authorization is* conjunctive*—most workflows require a* bundle *of credentials to proceed, and any missing credential blocks the entire action—and it is the regime under which Proposition S8 guarantees that the coalition game under union is superadditive, so that the grand coalition is optimal when all edges use union and disclosure is costless. Under cardinality, the valuation of the worked examples, only the weaker conclusion holds: enlarging a coalition under union never decreases its value, but may increase it by less than the sum of the parts. Nothing else in the paper depends on the regime.*

S4.8 Summary of Assumptions

The following table collects all simplifying assumptions and their effect on the framework:

#	Assumption	General (§§3–4)	Simplified (§§5–9)
A1	Credentials	Fuzzy $[0, 1]$, structured	Binary, atomic
A2	Scores	Embedded in credentials	Separate $s : \mathcal{A} \to \mathbb{R}^k$
A3	Graph	Directed, cyclic	Rooted tree
A4	Root type	Any	Human or agent (both discussed)
A5	Operators	Arbitrary	Four fundamentals
A6	Disclosure	Simultaneous	Sequential (root to leaves)
A7	Value function W	Monotone	Cardinality (monotone, not superadditive); superadditive only in Prop. S8

TABLE S1

Simplifying assumptions. The "General" column describes the framework as developed in Sections 3–4; the "Simplified" column describes the specialization adopted for the worked examples.

Remark S24 (Assumptions as a package)**.** *The seven assumptions are not independent in their effect. Their principal joint contribution is* tractability*: the tree topology (A3) together with sequential disclosure (A6) enables backward induction, reducing an exponential strategy space to a linear sequence of local decisions. The four fundamental operators (A5) on a tree guarantee that each edge's computation is self-contained—no multi-path confluence rules are needed. Binary credentials (A1) keep the coalition game finite and the Shapley computation exact. Credential complementarity (A7) is the one assumption the examples do* not *adopt: it is invoked only by Proposition S8, which makes the cooperative game superadditive under union. Relaxing any single assumption is straightforward; relaxing several simultaneously—for example, cyclic graphs with simultaneous disclosure and fuzzy credentials—would require substantially different solution techniques (see Sections 7 of the main text: "Beyond trees" and "Scaling to large networks").*

The general framework (Sections 3 of the main text–5 of the main text) and simplifying assumptions (A1–A7 above) are now in place. We apply them to four examples of increasing complexity, each introducing new operators, topologies, or game-theoretic phenomena. Example 1 uses intersection on a linear chain. Example 2 introduces mixed operators and reveals the destructive intermediary phenomenon. Example 3 adds cross-organizational structure and mechanism design. Example 4 combines all four operators in a six-actor stress test.

S5 Discussion of the Worked Examples

S5.1 The Destructive Intermediary

That an actor's presence can *reduce* what a coalition is authorized to do is, by this point in the paper, a theorem rather than a finding: Theorem 4.2 prices the effect exactly, Proposition 4.2 localizes it to intermediaries, and Theorem 4.1 shows that no topology avoids it. The worked examples supplied the first sightings—a human manager with $\Delta_h = -1$ already in the minimal mixed setting of Example 2 (Appendix S8), amplified to negative marginals in 12 of 16 re-attaching coalitions in Example 4 (Section 4.2 of the main text)—and classical delegation models cannot even represent the phenomenon. What merits discussion is not the numbers but three consequences.

First, this is an algebraic property of the *position and operator*, not of the actor type: nothing in equation (S3) mentions whether an actor is human or artificial. Any actor—human or AI—placed at the same intermediary node with the same intersection edge would produce the same destructive effect. The examples feature humans in this role because human-in-the-loop oversight is the most common design pattern, but the pathology is structural, not biological.

Second, the pathology is repairable. The worked examples use intersection edges without relevance filters (R); as the sensitivity analyses in Examples 3 and 4 (Appendices S8 and S9) demonstrate, introducing an appropriate R—for instance, replacing pure intersection with a filtered join that preserves the intermediary's oversight while passing through unique credentials—can substantially rebalance the coalition values and eliminate the negative marginal. The destructive intermediary is thus a *design flaw* that the framework diagnoses, not an inevitable consequence of including a human. Third, the credential propagation graph models authorization structure, not the operational workflow: an actor may be algebraically destructive yet operationally indispensable for regulatory or ethical reasons.

Our algebraic finding resonates with experimental evidence from outside the authorization domain. Terrucha et al. [S19] show that when humans program artificial delegates in collective-risk dilemmas, the resulting coalitions contribute more to collective goods but suffer from *precision errors* that the human principal introduces into the agent's algorithm—errors that cannot be corrected mid-game. Similarly, de Melo et al. [S20] find in behavioral experiments that humans cooperate *more* when acting through autonomous machines than when acting directly. Both results point in the same direction as our algebra: the human's direct participation can degrade collective outcomes relative to agent-mediated alternatives, whether through imprecise delegation or through destructive credential filtering.

S5.2 Operator Choice Dominates Credential Count

A second consequence of the structural theory is that an actor's value is driven by the *operator on its edge*, not by the size of its credential set: in the criterion (S3), what an actor's presence is worth depends on pass filters and unique coverage, and $|\sigma(a)|$ appears nowhere. The Shapley analysis of Example 2 (Appendix S8) makes the point numerically: three actors holding five, three, and four credentials receive the *same* Shapley value (2.00 each), because the union edge amplifies the smallest holder's unique credential into every coalition it joins, while the intersection edge forfeits the data agent's four unique credentials whenever the manager is present and charges the manager for the value it destroys.

This has a direct implication for mechanism design: in a network where operators are fixed, *an actor's compensation or priority should reflect its Shapley value, not its credential count*. A licensing body that grants more credentials does not necessarily create more value if those credentials enter the network through a restrictive operator.

S5.3 Strategic Distortion Amplifies Algebraic Effects

The algebra determines what credentials *could* propagate; the game theory determines which credentials *actually do*. When these diverge, the realized authorization differs from the algebraically feasible authorization.

In Example 2 (Appendix S8), the risk agent a_1 is the sole holder of c_7 (risk model certification). If disclosure of c_7 carries a cost exceeding its marginal benefit—say, revealing a proprietary risk model exposes competitive intelligence—then a_1's equilibrium strategy is to withhold it. The grand coalition's effective credential set drops from 6 to 5, and the human's marginal contribution to $\{a_1, a_2\}$ worsens from -1 to -2.

The strategic layer thus *amplifies* the algebraic effect: the intersection edge is already destructive, but under-disclosure makes it worse—an interaction invisible to either layer in isolation.

More generally, Proposition S11 shows that union edges create a free-rider problem: actors with overlapping credentials each prefer the other to bear the disclosure cost. This means that even in networks where the algebra permits full credential pooling, the equilibrium outcome may be a strict subset of the feasible set. The gap between algebraic feasibility and strategic equilibrium is a measurable quantity—and a target for mechanism design interventions.

Remark S25 (The algebra-strategy division of labor)**.** *A candid assessment: the credential algebra performs the primary analytical work in this paper. The counterintuitive findings—operator dominance, negative human marginals, the non-optimality of the grand*

coalition—are all algebraic results that hold under full disclosure. The game-theoretic layer's contribution is to show that these effects survive and intensify *under strategic behavior: the free-rider problem on union edges (Proposition S11), the amplification of destructive intersection through withholding, and the VCG mechanism for eliciting cross-organizational disclosure are all consequences of layering strategy on top of the algebra. The game theory does not* create *the headline findings—it confirms that strategic behavior cannot rescue the system from algebraic pathologies and may deepen them. Computing Bayesian equilibria under incomplete information, or analyzing repeated-game dynamics where trust evolves, would strengthen the strategic layer and is the most promising direction for future work.*

S6 Open Problems and Future Directions (Extended)

Several extensions arise naturally from the framework.

Beyond trees: The restriction to rooted trees (Assumption A3) is the one this paper relaxes furthest: Section 4.3 of the main text defines propagation on arbitrary directed graphs as the least solution of the propagation equations, and semantics, evaluation, the monotonicity dichotomy, and hardness all live there. What remains tree-bound is exact accounting: on graphs with reconvergent walks the destructive intermediary criterion degrades to two-sided bounds, and a closed form for the value lost through one filter and resupplied along another walk is open; target-directed synthesis beyond trees is Steiner-flavored and likewise open. Relaxing sequential disclosure (Assumption A6) replaces backward induction with normal-form games where mixed-strategy equilibria must be computed directly. Relaxing multiple assumptions simultaneously—cyclic graphs with simultaneous disclosure and fuzzy credentials—would likely require computational approximations rather than closed-form solutions.

Scaling to large networks: The exact Shapley computation in our examples is feasible because $n \leq 6$ (requiring $2^n = 64$ coalition evaluations). Real-world deployments with hundreds or thousands of agents will require approximate methods. Monte Carlo sampling of random permutations [S7] can estimate Shapley values in polynomial time with bounded error. For the credential propagation itself, message-passing algorithms on the actor graph—similar to those used in probabilistic graphical models—could compute effective credential sets without enumerating all coalitions. These numerical methods would form the computational backbone of any practical implementation of the framework.

Endogenous relevance filters: The relevance filter R on join edges is the framework's most consequential modeling choice and its most significant limitation. As the sensitivity analyses in Examples 3 and 4 demonstrate, small changes in R can shift results from conservative credential sharing to broad cross-organizational access, alter coalition values, and change Shapley rankings. Yet R is treated as exogenous throughout (Remark S3).

Endogenizing R is substantially harder than the disclosure game we analyze. It requires a *two-stage game*: in the first stage, actors negotiate or a regulator sets the relevance filters; in the second stage, the disclosure game proceeds as in our framework. The first stage is a mechanism design problem in its own right—actors have private information about which credentials they would prefer to filter (or to have filtered), and the filter choice affects both the algebraic structure and the subsequent strategic equilibrium. This two-stage structure creates circular dependencies: the optimal R depends on the disclosure equilibrium, which itself depends on R.

We do not attempt this extension here. In the domains we study (regulated finance, cross-organizational compliance), R is typically fixed by regulation or data-sharing agreements, making the exogenous assumption natural. But for domains where filters emerge from negotiation, endogenizing R is the most important open problem the framework raises: a two-stage game in which actors first agree on the filters and then play the disclosure game on the resulting edges.

Temporal dynamics: Credentials expire, are revoked, or are issued over time. The current framework is static: all credentials exist simultaneously. A dynamic extension would model credential evolution as a process on the actor graph, with operators applied at each time step to a changing credential landscape. The dynamics are especially acute in current agentic practice, where authorization is increasingly *sessional*: short-lived tokens are issued per subtask and agents are spawned per run, so the actor graph can change faster than a sampling estimate of the Shapley value converges. Instance-level auditing survives this regime—each configuration snapshot is evaluated in linear time—but stable value attribution over a drifting topology is an open problem. This connects naturally to the literature on dynamic mechanism design [S15]. On cyclic actor graphs a second dynamic phenomenon appears even with a fixed credential landscape. Rounds of re-evaluation—iterating the propagation map of Definition 3.11 from the state in which every actor holds its intrinsic set—converge to the least solution whenever every actor's incoming pass filters cover $\mathcal{C}$ (the start is then a sub-solution and the iteration is monotone), but can oscillate otherwise: on a two-cycle with one intersection and one union edge the iteration has period two, an authorization *flapping* in which two actors alternately grant and withdraw each other's credentials and no stationary answer exists. The least-solution semantics fixes the static answer but not the transient; characterizing the convergence of re-evaluation dynamics on cyclic graphs is open. The phenomenon has a close relative in interdomain routing, where the stable paths problem of Griffin, Shepherd, and Wilfong [S21] shows that locally consistent route selection can lack a stable solution and oscillate forever; whether an analogue of their dispute-wheel obstruction characterizes authorization flapping is a natural question.

Operator completeness: The four fundamental operators are the canonical instances of the pass-filtered class $\mathcal{P}$ (Definition 3.8), for which every structural and complexity result holds. Operators outside $\mathcal{P}$—for instance a gated pass-through $\otimes(X, Y) = Y$ if $X \neq \emptyset$ and $\emptyset$ otherwise, which is monotone and bounded but has no uniform child filter—fall outside the theory: the pass-bound lemma fails for them, and whether a weaker dichotomy survives is open. Characterizing

the largest operator class for which the dichotomy holds, and whether every monotone bounded operator with a uniform child filter is pass-filtered, are natural questions.

Fuzzy and probabilistic credentials: The binary credential model (Assumption A1) is a simplification. A fuzzy extension via the Zadeh [S18] operators (min for intersection, max for union) preserves the algebraic properties but introduces continuous-valued credential sets, changing the structure of the disclosure game from a discrete to a continuous-action space. A probabilistic extension, where credential possession is uncertain, would connect to Bayesian game theory and mechanism design under incomplete information. A third direction is *parameterized* credentials, as in modern cloud authorization where a right carries predicates over attributes rather than a bare token; flattening such predicates into atomic credentials can blow up $|\mathcal{C}|$ combinatorially. Our evaluation results are linear in the *size of the representation*, so the right generalization is a symbolic one—set operations over compactly represented predicate families—rather than naive flattening. Section S4.1 sharpens the point: an atom is an attestation, not an instantiation, so the blow-up arises only when credentials *entail* one another; that case—a distributive lattice of structured permissions in place of the powerset $\mathcal{P}(\mathcal{C})$, with pass filters as lattice elements—is the extension to pursue.

Implementation: The framework is theoretical, but it provides the algebraic theory that existing protocol implementations currently lack. The Authenticated Workflows protocol [S22] implements intersection-on-a-tree as a security primitive; our framework shows this is one special case of a richer operator algebra and identifies the game-theoretic incentives that the protocol does not address. The zero-trust identity framework of Huang et al. [S23] builds on DIDs and Verifiable Credentials to establish rich agent identities; our credential algebra specifies how those credentials should *compose* across actors, not merely how they are stored and verified. Extending these protocols to support union and join operators—and embedding strategic incentives into the protocol layer—is a concrete engineering challenge that our theory can guide. At deployment scale—networks of hundreds or thousands of agents—exact Shapley computation becomes infeasible, and the Monte Carlo and message-passing approximations discussed above ("Scaling to large networks") would form the computational layer beneath any such protocol implementation.

S7 Notation Reference

Symbol	Meaning
$\mathcal{A}$; $\mathcal{A}_H, \mathcal{A}_M$	Actor universe; human and machine actors
$\mathcal{C}$; $\mathcal{P}(\mathcal{C})$	Credential universe; its power set
$\sigma(a)$	Intrinsic credential set of actor a
$\hat{\sigma}(a)$, $\hat{\sigma}_S(a)$	Effective credential set of a (within coalition S)
$\hat{\sigma}_S$	Coalition effective set $\bigcup_{a\in S}\hat{\sigma}_S(a)$
$\mathcal{G}=(\mathcal{A},E)$; F_S	Actor graph; subgraph induced by S
$\otimes_{a,b}$; Ω	Credential operator on edge (a,b); operator assignment
$\cap, \cup, \ltimes_R, \rtimes_R$	The four fundamental operators; R a relevance filter
$\mathcal{P}$; (α,π)	Pass-filtered operators; parent and child filters
$\pi(\otimes)$	Pass filter of $\otimes$; preserving if and only if $\pi=\mathcal{C}$
O; O_S; $\mathrm{src}(F_S)$	Origins; origins of coalition S; sources of F_S
Θ_S; $\perp$	Propagation map on $\mathcal{P}(\mathcal{C})^S$; empty assignment
$\mathrm{val}(w)$	Value of a walk w
Φ	Multi-path aggregator (default: union)
$\delta_{a\to b}$; δ_i	Disclosed set on an edge; disclosed set of actor i
V_i, D_i	Value and disclosure-cost functions of player i
W; $v(S)$	Coalition valuation; characteristic function
$\Delta_x(S)$; ϕ_x	Marginal contribution; Shapley value
b_i, π_i; N, L; U_i	Filtered children of x and their filters; independent coverage and loss set; uniquely covered credentials

S8 Worked Examples 1–3

This appendix develops the first three worked examples in full, in escalating order of structural complexity: a linear three-actor intersection chain recovering classical delegation (Example 1), a three-actor branching coalition with mixed operators in which the destructive intermediary first appears (Example 2), and a cross-organizational investigation exercising the join operators and the VCG mechanism (Example 3). Example 4, which uses all four operators at depth two, is summarized in Section 4.2 of the main text and developed in full in Appendix S9.

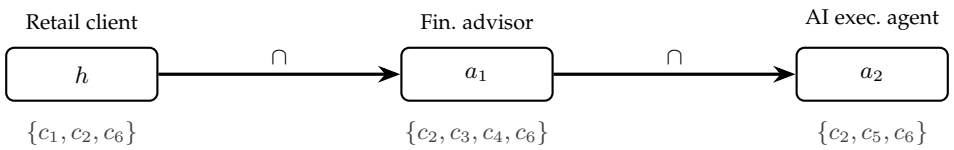
S8.1 Example 1: Trade Execution Chain

We begin with the simplest non-trivial case: a linear chain of three actors connected by intersection operators. This example recovers the classical delegation model and illustrates backward induction on a sequential disclosure game.

S8.1.1 Setup

Actors:

- h (human, retail client): holds account-level permissions.
- a_1 (human, financial advisor): holds advisory and trading authorizations.
- a_2 (machine, AI execution agent): holds exchange connectivity certifications and algorithmic trading licenses.



Credential universe: We define $\mathcal{C} = \{c_1, c_2, c_3, c_4, c_5, c_6\}$ with the following interpretation:

Credential	Meaning
c_1	KYC-verified identity
c_2	Equities trading authorization
c_3	Derivatives trading authorization
c_4	Fiduciary license
c_5	Exchange API connectivity certification
c_6	AML (anti-money-laundering) clearance

Intrinsic credential sets:

$$\sigma(h) = \{c_1, c_2, c_6\} \quad \text{(KYC, equities, AML)} \tag{S12}$$
$$\sigma(a_1) = \{c_2, c_3, c_4, c_6\} \quad \text{(equities, derivatives, fiduciary, AML)} \tag{S13}$$
$$\sigma(a_2) = \{c_2, c_5, c_6\} \quad \text{(equities, exchange API, AML)} \tag{S14}$$

Graph: The actor graph is a directed path $h \to a_1 \to a_2$, rooted at h. Both edges carry the intersection operator $\cap$.

S8.1.2 Credential Propagation

By Definition 3.9 of the main text (path propagation):

$$\hat{\sigma}(h) = \sigma(h) = \{c_1, c_2, c_6\} \tag{S15}$$
$$\hat{\sigma}(a_1) = \hat{\sigma}(h) \cap \sigma(a_1) = \{c_1, c_2, c_6\} \cap \{c_2, c_3, c_4, c_6\} = \{c_2, c_6\} \tag{S16}$$
$$\hat{\sigma}(a_2) = \hat{\sigma}(a_1) \cap \sigma(a_2) = \{c_2, c_6\} \cap \{c_2, c_5, c_6\} = \{c_2, c_6\} \tag{S17}$$

Remark S26. *The effective credential set at the terminal node a_2 is $\{c_2, c_6\}$—equities trading and AML clearance. This is exactly the intersection $\sigma(h) \cap \sigma(a_1) \cap \sigma(a_2) = \{c_2, c_6\}$, confirming Proposition 3.3. The AI execution agent can trade equities (authorized by all three actors) but* not *derivatives (c_3 is held only by the advisor) and* not *via its exchange API credential (c_5 is held only by the agent). Intersection enforces the classical principle: delegation can only narrow.*

The effect of credential complementarity: The analysis above is silent on the *value* of the credentials lost to intersection. Under cardinality ($W(S) = |S|$), losing c_1 (KYC) from the effective set costs exactly 1 unit—the same as losing any other credential. But authorization is rarely additive: in practice, credentials unlock value in *bundles*.

To see why this matters, define a superadditive value function that captures a natural authorization pattern: the bundle $\{c_1, c_2, c_6\}$ (KYC + equities trading + AML clearance) together enables *compliant retail trading*—a capability worth more than its parts. Concretely, let:

$$W_{\text{comp}}(S) = |S| + \begin{cases} 2 & \text{if } \{c_1, c_2, c_6\} \subseteq S \\ 0 & \text{otherwise} \end{cases} \tag{S18}$$

This W_{comp} is superadditive: the bundle bonus rewards credential combinations, not individual holdings.

Under W_{comp}, the values at each node change significantly:

Node	$W(S) = \lvert S \rvert$	$W_{\text{comp}}(S)$	Note
h (root)	3	5	Has the full bundle $\{c_1, c_2, c_6\}$
a_1 (advisor)	2	2	$\{c_2, c_6\}$: missing c_1, no bonus
a_2 (agent)	2	2	$\{c_2, c_6\}$: same effective set

The cost of intersection is no longer 1 unit per lost credential—it is 3 units at the first edge ($h \to a_1$), because losing c_1 destroys the entire bundle bonus. This is the amplification effect of credential complementarity: restrictive operators are more costly when credentials interact synergistically, and the "narrowing tax" of intersection falls disproportionately on credential bundles. A system designer choosing between intersection and union on an edge must account not only for which credentials are filtered, but for which *bundles* are broken.

S8.1.3 Sequential Disclosure Game

We now overlay the game-theoretic layer. Each actor strategically chooses which credentials to disclose, trading off the value of a successful trade against disclosure costs.

Scores: Let $s(h) = (0.9)$, $s(a_1) = (0.85)$, $s(a_2) = (0.95)$, representing reputation or reliability ratings on $[0, 1]$.

Value and cost functions: For each actor i, the value function is:

$$V_i(\hat{\sigma}) = \alpha_i \cdot |\hat{\sigma}| \cdot s(j) \tag{S19}$$

where j is the counterparty on the relevant edge and $\alpha_i > 0$ is a scaling constant. The disclosure cost is:

$$D_i(\delta_i) = \beta_i \cdot |\delta_i| \tag{S20}$$

where $\beta_i > 0$ is the per-credential disclosure cost.

Backward induction: Under sequential disclosure (root to leaves), we solve by backward induction from the terminal node a_2.

Step 1: Agent a_2's decision. Given the effective credentials $\hat{\sigma}(a_1)$ disclosed by the path so far, a_2 chooses $\delta_{a_2} \subseteq \sigma(a_2)$ to maximize:

$$u_{a_2} = V_{a_2}(\hat{\sigma}(a_1) \cap \delta_{a_2}) - D_{a_2}(\delta_{a_2}) \tag{S21}$$

Since intersection with $\hat{\sigma}(a_1)$ means that only credentials in $\hat{\sigma}(a_1) \cap \sigma(a_2)$ can contribute to value, the agent's optimal strategy is to disclose exactly $\delta^*_{a_2} = \hat{\sigma}(a_1) \cap \sigma(a_2)$—the credentials that will survive intersection—and withhold the rest. Disclosing additional credentials (e.g., c_5) incurs cost without increasing the intersection.

Step 2: Advisor a_1's decision. Anticipating a_2's best response, a_1 chooses $\delta_{a_1} \subseteq \sigma(a_1)$ to maximize:

$$u_{a_1} = V_{a_1}(\hat{\sigma}(h) \cap \delta_{a_1}) - D_{a_1}(\delta_{a_1}) \tag{S22}$$

By identical logic, a_1's optimal strategy is $\delta^*_{a_1} = \hat{\sigma}(h) \cap \sigma(a_1) = \{c_2, c_6\}$.

Step 3: Client h's decision. As root, h discloses first. Since h's effective credentials are simply $\sigma(h)$, and every credential h withholds reduces the downstream intersection, h's optimal strategy depends on the trade-off between α_h and β_h. If α_h/β_h is sufficiently large (the trade is valuable relative to disclosure cost), h discloses $\delta^*_h = \sigma(h)$ in full.

Proposition S10 (SPE of the Linear Chain)**.** *Under the value and cost functions of equations* (S19)–(S20)*, the subgame perfect equilibrium of the sequential disclosure game on the linear chain $h \to a_1 \to a_2$ with intersection operators has each actor disclosing exactly the credentials that survive intersection with the accumulated path credentials. The equilibrium effective credential set at a_2 equals the full-disclosure intersection $\sigma(h) \cap \sigma(a_1) \cap \sigma(a_2)$, provided the value-to-cost ratio α_i/β_i exceeds 1 for all actors.*

Proof. By backward induction. At the last stage, a_2 observes the accumulated credentials $\hat{\sigma}(a_1)$ and chooses which of its own credentials to disclose. Under intersection, disclosing $c \in \sigma(a_2)$ adds c to the effective set only if $c \in \hat{\sigma}(a_1)$; otherwise it has no effect. The cost of disclosing c is β_2, and the benefit is α_2 if c survives intersection, zero otherwise. When $\alpha_2 > \beta_2$, a_2's best response is to disclose exactly $\sigma(a_2) \cap \hat{\sigma}(a_1)$. The same argument applies at each preceding stage: a_1 discloses $\sigma(a_1) \cap \hat{\sigma}(h)$, and h discloses all of $\sigma(h)$ (since the root has no upstream filter). The resulting effective set is the full intersection chain. □

Remark S27. *Under intersection, the disclosure game has a clean structure: no actor has an incentive to over-disclose (wasteful) or under-disclose (reduces value). This well-behavedness is specific to intersection. Under union, as we see in the next example, the incentive structure changes fundamentally.*

S8.2 Example 2: Portfolio Analysis Coalition

The second example introduces branching, mixed operators, and coalition formation. A human portfolio manager collaborates with two AI agents whose credentials complement the manager's own.

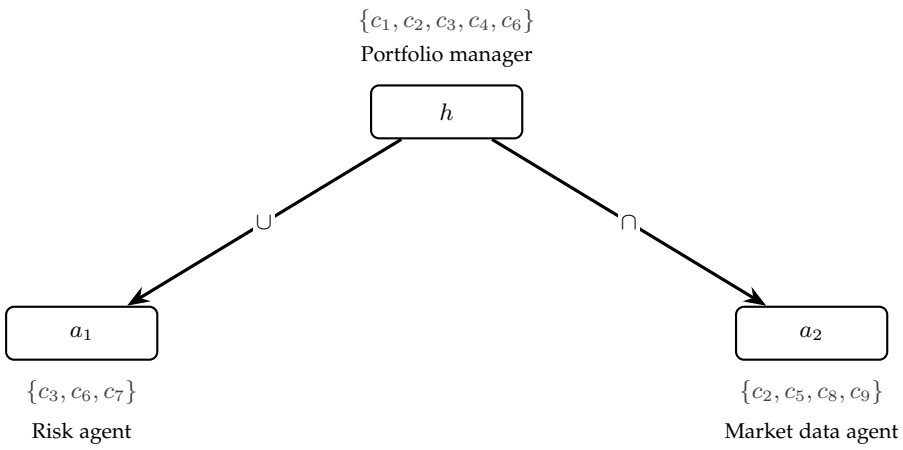
S8.2.1 Setup

Actors:

- h (human, portfolio manager): holds advisory authorizations and client relationship credentials.
- a_1 (machine, AI risk analysis agent): holds quantitative risk certifications.
- a_2 (machine, AI market data agent): holds real-time exchange data access and alternative data licenses.

Credential universe: We extend the universe to $\mathcal{C} = \{c_1, \ldots, c_9\}$:

Credential	Meaning
c_1	KYC-verified identity
c_2	Equities trading authorization
c_3	Derivatives trading authorization
c_4	Fiduciary license
c_5	Real-time exchange data access
c_6	AML clearance
c_7	Quantitative risk model certification (Basel III)
c_8	Alternative data license (satellite, sentiment)
c_9	Alternative trading venue access



Intrinsic credential sets:

$$\sigma(h) = \{c_1, c_2, c_3, c_4, c_6\} \quad \text{(KYC, equities, derivatives, fiduciary, AML)} \tag{S23}$$
$$\sigma(a_1) = \{c_3, c_6, c_7\} \quad \text{(derivatives, AML, risk model cert)} \tag{S24}$$
$$\sigma(a_2) = \{c_2, c_5, c_8, c_9\} \quad \text{(equities, exchange data, alt data, venue access)} \tag{S25}$$

Graph: The actor graph is a rooted tree with h at the root and two children: $h \to a_1$ and $h \to a_2$. The edge $h \to a_1$ carries the *union* operator $\cup$ (the risk agent's quantitative certifications complement the manager's), while the edge $h \to a_2$ carries *intersection* $\cap$ (a compliance restriction: the desk accepts from the data agent only credentials the manager independently holds).

S8.2.2 Credential Propagation

$$\hat{\sigma}(h) = \sigma(h) = \{c_1, c_2, c_3, c_4, c_6\} \tag{S26}$$
$$\hat{\sigma}(a_1) = \hat{\sigma}(h) \cup \sigma(a_1) = \{c_1, c_2, c_3, c_4, c_6\} \cup \{c_3, c_6, c_7\} = \{c_1, c_2, c_3, c_4, c_6, c_7\} \tag{S27}$$
$$\hat{\sigma}(a_2) = \hat{\sigma}(h) \cap \sigma(a_2) = \{c_1, c_2, c_3, c_4, c_6\} \cap \{c_2, c_5, c_8, c_9\} = \{c_2\} \tag{S28}$$

Remark S28. *The asymmetry is striking. Through the union edge, the risk agent* expands *the effective credential set: c_7 (Basel III risk certification) becomes available even though the portfolio manager does not hold it. This is impossible under classical delegation. Through the intersection edge, the market data agent's unique credentials (c_5, c_8, c_9) are lost because the manager does not hold them. The operator choice on each edge encodes a policy decision about how much trust the collaboration places in each agent's independent credentials.*

S8.2.3 Coalition Formation and Shapley Value

We now analyze the credential coalition game (Definition 5.4 of the main text) over this three-actor network.

Global value function: Let $W(\hat{\sigma}_S) = |\hat{\sigma}_S|$ under shared-outcome evaluation (Definition 3.11 of the main text)—the coalition's value is the size of the effective credential set its members assemble together. This is a monotone, submodular function.

Coalition values: Under optimal disclosure with zero cost (to isolate the coalitional structure), the characteristic function values are:

Coalition S	$v(S)$	Effective credentials
$\{h\}$	5	$\{c_1, c_2, c_3, c_4, c_6\}$
$\{a_1\}$	3	$\{c_3, c_6, c_7\}$
$\{a_2\}$	4	$\{c_2, c_5, c_8, c_9\}$
$\{h, a_1\}$	6	$\{c_1, c_2, c_3, c_4, c_6, c_7\}$ (union)
$\{h, a_2\}$	5	$\{c_1, c_2, c_3, c_4, c_6\}$ (h's set; a_2 retains only c_2)
$\{a_1, a_2\}$	7	$\{c_2, c_3, c_5, c_6, c_7, c_8, c_9\}$ (no edge; intrinsic sets, disjoint)
$\{h, a_1, a_2\}$	6	$\{c_1, c_2, c_3, c_4, c_6, c_7\}$ (grand coalition)

Remark S29. *The coalition $\{a_1, a_2\}$ has no edge in the tree (they are siblings), so each actor is a component root and contributes its intrinsic set (Definition 3.11 of the main text); the two sets are disjoint, hence $v(\{a_1, a_2\}) = 3 + 4 = 7$. Note that $v(\{a_1, a_2\}) = 7 > 6 = v(\{h, a_1, a_2\})$: the grand coalition is worth* less *than two disconnected agents, because adding h introduces the intersection edge to a_2 which destroys credentials. This is a striking consequence of mixing operators.*

An important caveat: the actor graph models credential propagation*, not the operational workflow itself. Saying that removing h improves the coalition's credential value does not mean the human should be removed from the process—the operation may require*

human oversight for legal, regulatory, or ethical reasons that the algebra does not capture. Rather, the result tells a system architect that the human's credential edge *(intersection) is the problem: redesigning the operator—for instance, replacing intersection with a left join that preserves the human's oversight role while filtering less aggressively—could retain the human in the workflow without the algebraic penalty.*

Remark S30. *The coalition $\{h, a_2\}$ is notable: despite $5 + 4 = 9$ intrinsic credentials between them, the collaboration is worth exactly $v(\{h\}) = 5$—the intersection edge reduces a_2's four credentials to $\{c_2\}$, which h already holds, so the data agent brings nothing that survives ($\Delta_{a_2}(\{h\}) = 0$), while h's own set is untouched ($\Delta_h(\{a_2\}) = +1$). Under at-target evaluation at a_2 the same edge is outright destructive: $|\{c_2\}| - |\sigma(a_2)| = -3$. The coalition $\{h, a_1\}$ under union fares better: $v(\{h, a_1\}) = 6 > \max(v(\{h\}), v(\{a_1\})) = 5$, so the coalition is monotone (enlarging helps). However, $6 < 5 + 3 = 8$, so even under union the coalition is not superadditive at the v-level when $W(S) = |S|$ (cardinality), because overlapping credentials (c_3, c_6 held by both) contribute only once to the union. Superadditivity of v requires superadditivity of W (credential complementarity, Assumption A7), which is stronger than cardinality—it holds when credentials unlock value* synergistically *rather than independently.*

Shapley values: Computing the Shapley value (Definition 5.6 of the main text) for each actor by averaging marginal contributions over all $3! = 6$ orderings:

Actor	ϕ_i	Interpretation
h	$\frac{12}{6} = 2.00$	Broadest holdings, but a negative marginal to $\{a_1, a_2\}$ through the intersection edge
a_1	$\frac{12}{6} = 2.00$	Fewest holdings, but its unique c_7 expands every coalition through the union edge
a_2	$\frac{12}{6} = 2.00$	Four unique credentials, all lost whenever h is present; contributes only without h

The full computation over all six orderings is given in Appendix S8.2.5.

Remark S31. *The Shapley values reveal several notable features. All three actors receive the* same *attribution, 2.00, although their holdings (five, three, and four credentials) and roles differ sharply: the manager's breadth is offset by the value its intersection edge destroys, the risk agent's three credentials are amplified by the union edge that carries its unique c_7 into every coalition, and the data agent's four unique credentials count only in coalitions from which the manager is absent ($\Delta_{a_2}(\{h\}) = \Delta_{a_2}(\{h, a_1\}) = 0$). Attribution nets destruction against contribution. Most strikingly, h's marginal contribution to $\{a_1, a_2\}$ is* negative*: $\Delta_h(\{a_1, a_2\}) = 6-7 = -1$—exactly the criterion of Theorem 4.2: gain $\{c_1, c_4\}$, loss $\{c_5, c_8, c_9\}$ (the credentials only a_2 covers and the filter removes), $2-3 = -1$. Adding the human manager to the two-agent coalition makes it* worse*, because h's presence activates the intersection edge to a_2, destroying credentials that were available when the agents operated independently. This demonstrates that in mixed-operator networks, adding actors can reduce coalition value—a phenomenon impossible under pure union.*

S8.2.4 Strategic Disclosure under Union

The disclosure game on the union edge $h \to a_1$ differs qualitatively from the intersection case of Example 1.

Under union, every credential that a_1 discloses is *added* to the effective set, regardless of what h holds. This creates a free-rider problem: the portfolio manager benefits from a_1's disclosure without needing to reciprocate with credentials that a_1 values. If disclosure costs are positive, a prisoners'-dilemma dynamic can emerge where both actors under-disclose, yielding an equilibrium effective set smaller than the full-disclosure union.

Proposition S11 (Under-disclosure under Union)**.** *In the disclosure game on a union edge (a, b) with $D_i(\delta) = \beta \cdot |\delta|$ for $\beta > 0$ and $V_i(\hat{\sigma}) = \alpha \cdot |\hat{\sigma}|$, a pure-strategy Nash equilibrium exists in which each actor discloses only the credentials that are unique to itself (i.e., not held by the other), provided $\alpha > \beta$. Credentials held by both actors are withheld by at least one party.*

Proof. If $c \in \sigma(a) \cap \sigma(b)$, then c appears in the union regardless of whether a discloses it, provided b discloses it (and vice versa). Each actor prefers the other to bear the cost. For shared credentials, the game on each credential is a game of chicken with two pure equilibria (one discloses, the other free-rides) and a mixed equilibrium. For unique credentials $c \in \sigma(a) \setminus \sigma(b)$, actor a is the sole provider; disclosure is optimal whenever $\alpha > \beta$. □

Strategic distortion in Example 2: Proposition S11 has a concrete and striking consequence in this example. The risk agent a_1 is the sole holder of c_7 (risk model certification). Disclosing c_7 may carry a non-trivial cost d_7—for instance, revealing a proprietary risk model exposes competitive intelligence. If $d_7 > \alpha$ (the marginal benefit of one additional credential in the effective set), a_1's best response is to *withhold* c_7.

Under this strategic equilibrium, the effective credential set of the grand coalition drops from 6 to 5, since c_7 was the unique credential that a_1 contributed via union. The coalition values shift: $v(\{h, a_1\}) = 5$ (no longer exceeding $v(\{h\})$) and $v(\{h, a_1, a_2\}) = 5$. Against the seven credentials the two agents hold on their own, the grand coalition now assembles five: one is destroyed by the human's intersection edge, one is withheld strategically. The algebra's loss and the game's loss add.

This is precisely the interaction the framework is designed to capture: the algebra determines *what credentials could propagate*, while the game theory determines *which credentials actually do*. When these diverge, the realized authorization differs from the algebraically feasible authorization. We develop this observation further in Section 7 of the main text.

S8.2.5 Shapley Value Computation for Example 2

We compute the Shapley values for the three-actor credential coalition game of Section S8.2. Recall the characteristic function:

Coalition	$v(S)$
$\{h\}$	5
$\{a_1\}$	3
$\{a_2\}$	4
$\{h, a_1\}$	6
$\{h, a_2\}$	5
$\{a_1, a_2\}$	7
$\{h, a_1, a_2\}$	6

The value $v(\{a_1, a_2\}) = 7$ is the size of $\sigma(a_1) \cup \sigma(a_2)$: the siblings have no connecting edge, each is a component root contributing its intrinsic set, and the two sets are disjoint. The value $v(\{h, a_2\}) = 5$ is $|\sigma(h) \cup (\sigma(h) \cap \sigma(a_2))| = |\sigma(h)|$.

The Shapley value for actor i averages the marginal contribution $\Delta_i(S) = v(S \cup \{i\}) - v(S)$ over all subsets:

$$\phi_i = \sum_{S \subseteq \mathcal{A} \setminus \{i\}} \frac{|S|!\,(|\mathcal{A}| - |S| - 1)!}{|\mathcal{A}|!}\, \Delta_i(S) \tag{S29}$$

Shapley value for h::

$$\begin{aligned}
\Delta_h(\emptyset) &= v(\{h\}) - v(\emptyset) = 5 - 0 = 5 && \text{weight } \tfrac{0!\,2!}{3!} = \tfrac{2}{6} \\
\Delta_h(\{a_1\}) &= v(\{h, a_1\}) - v(\{a_1\}) = 6 - 3 = 3 && \text{weight } \tfrac{1!\,1!}{3!} = \tfrac{1}{6} \\
\Delta_h(\{a_2\}) &= v(\{h, a_2\}) - v(\{a_2\}) = 5 - 4 = 1 && \text{weight } \tfrac{1!\,1!}{3!} = \tfrac{1}{6} \\
\Delta_h(\{a_1, a_2\}) &= v(\{h, a_1, a_2\}) - v(\{a_1, a_2\}) = 6 - 7 = -1 && \text{weight } \tfrac{2!\,0!}{3!} = \tfrac{2}{6}
\end{aligned}$$

$$\phi_h = \tfrac{2}{6}(5) + \tfrac{1}{6}(3) + \tfrac{1}{6}(1) + \tfrac{2}{6}(-1) = \tfrac{10+3+1-2}{6} = \tfrac{12}{6} = 2.00$$

Shapley value for a_1::

$$\begin{aligned}
\Delta_{a_1}(\emptyset) &= v(\{a_1\}) - 0 = 3 && \text{weight } \tfrac{2}{6} \\
\Delta_{a_1}(\{h\}) &= v(\{h, a_1\}) - v(\{h\}) = 6 - 5 = 1 && \text{weight } \tfrac{1}{6} \\
\Delta_{a_1}(\{a_2\}) &= v(\{a_1, a_2\}) - v(\{a_2\}) = 7 - 4 = 3 && \text{weight } \tfrac{1}{6} \\
\Delta_{a_1}(\{h, a_2\}) &= v(\{h, a_1, a_2\}) - v(\{h, a_2\}) = 6 - 5 = 1 && \text{weight } \tfrac{2}{6}
\end{aligned}$$

$$\phi_{a_1} = \tfrac{2}{6}(3) + \tfrac{1}{6}(1) + \tfrac{1}{6}(3) + \tfrac{2}{6}(1) = \tfrac{6+1+3+2}{6} = \tfrac{12}{6} = 2.00$$

Shapley value for a_2::

$$\begin{aligned}
\Delta_{a_2}(\emptyset) &= v(\{a_2\}) - 0 = 4 && \text{weight } \tfrac{2}{6} \\
\Delta_{a_2}(\{h\}) &= v(\{h, a_2\}) - v(\{h\}) = 5 - 5 = 0 && \text{weight } \tfrac{1}{6} \\
\Delta_{a_2}(\{a_1\}) &= v(\{a_1, a_2\}) - v(\{a_1\}) = 7 - 3 = 4 && \text{weight } \tfrac{1}{6} \\
\Delta_{a_2}(\{h, a_1\}) &= v(\{h, a_1, a_2\}) - v(\{h, a_1\}) = 6 - 6 = 0 && \text{weight } \tfrac{2}{6}
\end{aligned}$$

$$\phi_{a_2} = \tfrac{2}{6}(4) + \tfrac{1}{6}(0) + \tfrac{1}{6}(4) + \tfrac{2}{6}(0) = \tfrac{8+0+4+0}{6} = \tfrac{12}{6} = 2.00$$

Verification: $\phi_h + \phi_{a_1} + \phi_{a_2} = \tfrac{12+12+12}{6} = \tfrac{36}{6} = 6 = v(\{h, a_1, a_2\})$. ✓

Remark S32. *Several features of this computation are noteworthy:*

*(i) All three actors have the same Shapley value (*2.00*) despite holding five, three, and four credentials: attribution nets each actor's contributions against the value its edges destroy or forfeit.*

(ii) The risk agent a_1 matches the manager with the fewest intrinsic credentials because the union operator on its edge carries its unique c_7 into every coalition it joins.

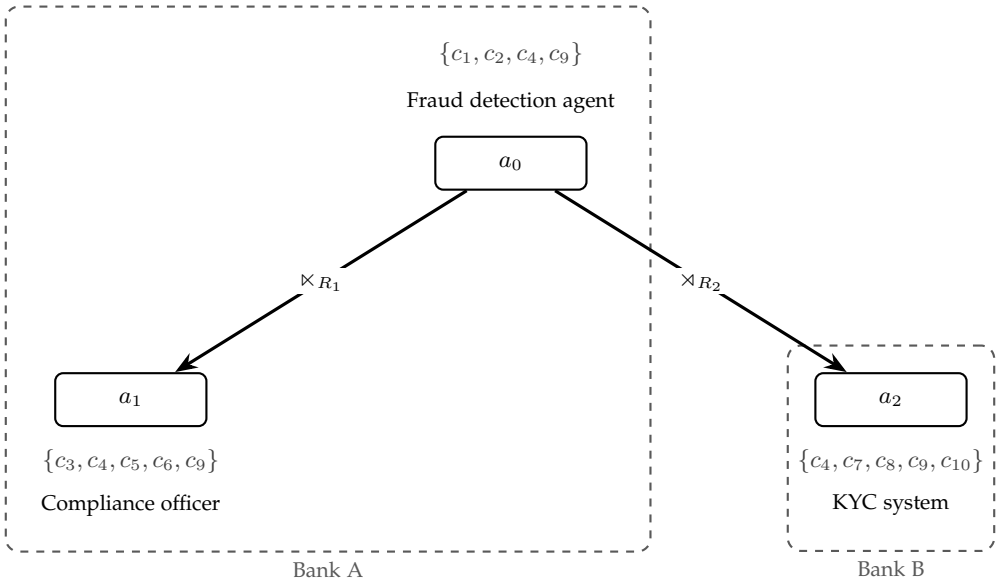
(iii) h's marginal contribution to $\{a_1, a_2\}$ is negative *(-1): adding h to two independently operating agents makes the coalition worse by activating the destructive intersection edge to a_2.*

(iv) a_2's marginal contribution to any coalition containing h is zero: the intersection edge reduces its four credentials to $\{c_2\}$, which h already holds. Under at-target evaluation at a_2 the same edge is destructive ($1 - 4 = -3$).

These results illustrate that in mixed-operator networks, the Shapley value captures not only positive contributions but also the damage that an actor's presence can inflict through restrictive operators.

S8.3 Example 3: Cross-Organizational Fraud Investigation

The third example introduces an agent-rooted workflow, asymmetric join operators, multiple organizations, and mechanism design, showcasing features that the classical delegation model cannot express.



S8.3.1 Setup

Scenario: An autonomous fraud detection agent at Bank A identifies a suspicious transaction pattern and initiates an investigation that requires credentials from both Bank A and Bank B. No human initiated the workflow.

Actors:

- a_0 (machine, fraud detection agent at Bank A): the root. Holds fraud detection certifications and Bank A internal access.
- a_1 (human, compliance officer at Bank A): holds regulatory reporting authority and investigation credentials.
- a_2 (machine, KYC verification system at Bank B): holds Bank B's customer identity verification credentials and sanctions screening certifications.

Credential universe: $\mathcal{C} = \{c_1, \ldots, c_{10}\}$:

Credential	Meaning
c_1	Fraud detection certification
c_2	Bank A transaction log access
c_3	Suspicious Activity Report (SAR) filing authority
c_4	AML clearance
c_5	Bank A internal investigation authorization
c_6	Regulatory reporting authority (SEC/FINRA)
c_7	Bank B customer identity records access
c_8	Sanctions screening certification (Office of Foreign Assets Control)
c_9	Cross-organizational data sharing agreement
c_{10}	Bank B internal audit trail access

Intrinsic credential sets:

$$\sigma(a_0) = \{c_1, c_2, c_4, c_9\} \quad \text{(fraud cert, logs, AML, data sharing)} \tag{S30}$$
$$\sigma(a_1) = \{c_3, c_4, c_5, c_6, c_9\} \quad \text{(SAR, AML, investigation, regulatory, data sharing)} \tag{S31}$$
$$\sigma(a_2) = \{c_4, c_7, c_8, c_9, c_{10}\} \quad \text{(AML, customer IDs, sanctions, data sharing, audit)} \tag{S32}$$

Graph: Rooted tree with agent root a_0, two children: $a_0 \to a_1$ and $a_0 \to a_2$. The edge $a_0 \to a_1$ carries the *left join* $\ltimes$ with relevance filter $R_1 = \{c_3, c_5, c_6\}$ (the fraud agent's credentials are primary; the compliance officer contributes only investigation-relevant credentials). The edge $a_0 \to a_2$ carries the *right join* $\rtimes$ with relevance filter $R_2 = \{c_1, c_4, c_9\}$ (Bank B's credentials are primary; Bank A contributes only credentials relevant to cross-organizational cooperation).

Dashed boxes indicate organizational boundaries; edges crossing these boundaries represent cross-organizational credential flows governed by relevance filters.

S8.3.2 Credential Propagation

$$\hat{\sigma}(a_0) = \sigma(a_0) = \{c_1, c_2, c_4, c_9\} \tag{S33}$$

$$\begin{aligned}\hat{\sigma}(a_1) &= \hat{\sigma}(a_0) \ltimes \sigma(a_1) \\ &= \hat{\sigma}(a_0) \cup \big(\sigma(a_1) \cap R_1\big) \\ &= \{c_1, c_2, c_4, c_9\} \cup \big(\{c_3, c_4, c_5, c_6, c_9\} \cap \{c_3, c_5, c_6\}\big) \\ &= \{c_1, c_2, c_4, c_9\} \cup \{c_3, c_5, c_6\} \\ &= \{c_1, c_2, c_3, c_4, c_5, c_6, c_9\}\end{aligned} \tag{S34}$$

$$\begin{aligned}\hat{\sigma}(a_2) &= \hat{\sigma}(a_0) \rtimes \sigma(a_2) \\ &= \sigma(a_2) \cup \big(\hat{\sigma}(a_0) \cap R_2\big) \\ &= \{c_4, c_7, c_8, c_9, c_{10}\} \cup \big(\{c_1, c_2, c_4, c_9\} \cap \{c_1, c_4, c_9\}\big) \\ &= \{c_4, c_7, c_8, c_9, c_{10}\} \cup \{c_1, c_4, c_9\} \\ &= \{c_1, c_4, c_7, c_8, c_9, c_{10}\}\end{aligned} \tag{S35}$$

Remark S33 (Agent-rooted workflow)**.** *The fraud detection agent a_0 initiated this workflow autonomously. Under classical delegation, this scenario is undefined—there is no human principal from whom to derive authority. In our framework, a_0's intrinsic credentials (c_1: fraud certification, c_2: log access, c_4: AML, c_9: data sharing agreement) provide the initial authorization seed, and the algebra proceeds identically to a human-rooted tree.*

Remark S34 (Asymmetric joins in action)**.** *The left join on $a_0 \to a_1$ keeps the fraud agent's credentials primary and selectively admits the compliance officer's investigation credentials (c_3, c_5, c_6). The officer's AML clearance (c_4) and data sharing credential (c_9) are filtered out by R_1 because they are redundant with a_0's own credentials. The right join on $a_0 \to a_2$ reverses the asymmetry: Bank B's credentials are primary, and Bank A contributes only cooperation-relevant credentials (c_1, c_4, c_9) through the filter R_2. Bank A's internal log access (c_2) is excluded—Bank B should not inherit access to Bank A's transaction logs.*

Sensitivity to the relevance filter: Since the relevance filter R determines which credentials pass through a join, the effective credential sets—and hence the coalition values—are directly sensitive to R. We illustrate this by varying R_2 on the cross-organizational edge $a_0 \to a_2$.

R_2 variant	$\hat{\sigma}(a_2)$	$\lvert\hat{\sigma}(a_2)\rvert$
Restrictive: $\{c_9\}$	$\{c_4, c_7, c_8, c_9, c_{10}\}$	5
Baseline: $\{c_1, c_4, c_9\}$	$\{c_1, c_4, c_7, c_8, c_9, c_{10}\}$	6
Permissive: $\{c_1, c_2, c_4, c_9\}$	$\{c_1, c_2, c_4, c_7, c_8, c_9, c_{10}\}$	7

Under the restrictive filter, Bank B receives only the data-sharing agreement (c_9) from Bank A—the minimum needed for cross-organizational cooperation. Under the permissive filter, Bank B also inherits Bank A's internal log access (c_2), which the baseline filter deliberately excluded.

The effect is not merely quantitative. Adding c_2 (transaction log access) to R_2 changes the *qualitative* nature of the collaboration: Bank B's KYC system would gain visibility into Bank A's internal transactions, a significant privacy and competitive concession. Removing c_1 (fraud certification) from R_2 means Bank B cannot verify that Bank A's agent is a certified fraud investigator, weakening the trust basis for the collaboration.

This sensitivity underscores the importance of the relevance filter as a policy instrument. Small changes in R can shift results from conservative information sharing to broad cross-organizational access. In regulated domains, R is typically fixed by data-sharing agreements or regulatory mandate (our exogenous assumption), precisely because its effect on the credential algebra is so consequential.

S8.3.3 Mechanism Design for Cross-Organizational Disclosure

The key challenge in this example is that Bank A and Bank B are *competing institutions*. Bank B has no intrinsic incentive to disclose its customer identity records (c_7) or sanctions screening results (c_8) to Bank A's fraud agent. A mechanism is needed to elicit truthful disclosure.

The principal: The fraud detection agent a_0 (or, more precisely, the orchestration layer operating on a_0's behalf) acts as the mechanism designer.

Types: Each actor's type is its true credential set: $\theta_{a_1} = \sigma(a_1)$, $\theta_{a_2} = \sigma(a_2)$. These are private information.

Allocation rule: The allocation rule f computes the effective credential sets via the join operators as defined above, using the *reported* credential sets rather than the true ones.

Transfer rule: Following the VCG construction (Proposition S9), the transfer to actor a_i is:

$$t_{a_i}(\boldsymbol{\theta}) = \sum_{j \neq i} V_j\big(f(\boldsymbol{\theta})\big) - \sum_{j \neq i} V_j\big(f^{-i}(\boldsymbol{\theta}_{-i})\big) \tag{S36}$$

Concrete computation: The key modeling choice is the value function. In Examples 1 and 2, each actor's value depended only on its own effective credentials. But a fraud investigation is a *shared-outcome* task: the investigation succeeds or fails based on the combined evidence assembled across all actors, not on any individual actor's slice.

We therefore define a shared investigation value:

$$W_{\text{inv}}(\hat{\boldsymbol{\sigma}}) = |\hat{\sigma}(a_0) \cup \hat{\sigma}(a_1) \cup \hat{\sigma}(a_2)| \tag{S37}$$

Each actor's payoff is proportional to the investigation's overall capability: $V_i(\hat{\boldsymbol{\sigma}}) = W_{\text{inv}}(\hat{\boldsymbol{\sigma}})$ for all i. This captures the reality that all participants benefit from a successful investigation (through regulatory compliance, reputational protection, and fraud loss prevention).

Under the baseline credential sets:

$$\begin{aligned}\hat{\sigma}(a_0) \cup \hat{\sigma}(a_1) \cup \hat{\sigma}(a_2) &= \{c_1, c_2, c_4, c_9\} \cup \{c_1, c_2, c_3, c_4, c_5, c_6, c_9\} \cup \{c_1, c_4, c_7, c_8, c_9, c_{10}\} \\ &= \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}\}\end{aligned}$$

So $W_{\text{inv}} = 10$ (the full credential universe is covered).

Without a_2: The investigation proceeds with only Bank A actors:

$$\hat{\sigma}(a_0) \cup \hat{\sigma}(a_1) = \{c_1, c_2, c_3, c_4, c_5, c_6, c_9\},$$

so $W_{\text{inv}}^{-a_2} = 7$. Each remaining actor values this at 7.

VCG transfer for a_2: The externality a_2 imposes on others is the improvement in their joint value:

$$t_{a_2} = \sum_{j \neq a_2} \big[V_j(\text{with } a_2) - V_j(\text{without } a_2)\big] = 2 \times (10 - 7) = 6 \tag{S38}$$

Bank B's KYC system receives a VCG transfer of 6 units, reflecting the value of the three credentials (c_7, c_8, c_{10}) that only Bank B contributes to the investigation.

Without a_1: $\hat{\sigma}(a_0) \cup \hat{\sigma}(a_2) = \{c_1, c_2, c_4, c_7, c_8, c_9, c_{10}\}$, so $W_{\text{inv}}^{-a_1} = 7$. The VCG transfer for a_1:

$$t_{a_1} = 2 \times (10 - 7) = 6 \tag{S39}$$

The compliance officer also receives a transfer of 6 units, for the three credentials (c_3, c_5, c_6) unique to the Bank A investigation branch.

Remark S35 (VCG transfers under shared outcomes)**.** *The non-zero transfers arise because the shared-outcome value function creates interdependence between the branches: every actor benefits from every other actor's credentials, even on a tree. This is the natural model for collaborative investigations, joint audits, and any multi-party task where success depends on the* union *of evidence.*

The transfers have a natural interpretation: each actor is compensated in proportion to the unique *credentials it brings to the investigation. Since a_1 and a_2 each contribute exactly 3 unique credentials, their transfers are equal—despite operating at different banks, holding different credential types, and connecting through different operators. The VCG mechanism treats them as equally valuable to the investigation, which is the correct assessment under $W = |S|$.*

Under the superadditive W_{comp} from Example 1, transfers would be higher for actors whose unique credentials complete valuable bundles, providing stronger incentives for precisely those disclosures that are most consequential.

Remark S36 (Comparison with the zero-transfer case)**.** *If each actor's value depended only on its* own *effective credentials ($V_i(\hat{\boldsymbol{\sigma}}) = |\hat{\sigma}(a_i)|$, the individual-outcome model), the VCG transfer for a_2 would be zero: Bank B's participation does not change the effective credentials of Bank A's actors on a tree, since the branches are structurally independent. The choice between shared-outcome and individual-outcome value functions is therefore a consequential modeling decision. In domains where actors share the fruits of collaboration (investigations, audits, joint ventures), the shared-outcome model is appropriate and the VCG mechanism produces meaningful incentives.*

S8.3.4 Regulatory Constraints as Operator Selection

A distinctive feature of this example is that the choice of operators and relevance filters is not freely made by the actors but is *imposed by regulation*.

Anti-money-laundering regulations require that cross-organizational credential sharing be minimized to what is necessary for the investigation (data minimization principle). This maps directly to the relevance filters R_1 and R_2: the filters are set by compliance policy, not by strategic choice. The use of *joins* rather than *union* on cross-organizational edges reflects the regulatory requirement that neither bank should gain unrestricted access to the other's credential space.

Remark S37. *This illustrates a general pattern: in regulated environments, the game-theoretic layer operates* within constraints set by the algebra. *The operators and relevance filters encode regulatory policy; the game theory governs strategic behavior within those constraints. The framework thus separates two concerns:* what is permitted *(algebra) from* what is incentivized *(game theory).*

S9 EXAMPLE 4: FULL DEVELOPMENT

This appendix contains the complete development of Example 4 (summarized in Section 4.2 of the main text): the full setup with diagram, credential propagation, the coalition game, and the Shapley analysis. The exhaustive 63-coalition computation follows in Appendix S10.

The fourth example combines all elements of the framework in a single scenario: all four fundamental operators, six actors across three organizations, depth-two propagation, shared-outcome mechanism design, and both human and AI actors at every level of the tree.

S9.0.1 Setup

Scenario: A regulator mandates a coordinated stress test across two banks. An autonomous stress test coordinator agent assembles credentials from risk, compliance, and data actors at both institutions. No single actor holds sufficient credentials to conduct the stress test alone.

Actors:

- a_0 (machine, stress test coordinator): regulator-appointed agent that initiates and orchestrates the test. The root.
- h_1 (human, Bank A risk officer): manages Bank A's risk assessment function.
- a_1 (machine, Bank A risk model): quantitative risk engine.
- a_3 (machine, Bank A historical data repository): holds backtesting data and historical loss records.
- h_2 (human, Bank B compliance officer): manages Bank B's regulatory reporting.
- a_2 (machine, Bank B data system): holds Bank B's customer exposure data and sanctions screening.

Credential universe: $\mathcal{C} = \{c_1, \ldots, c_{12}\}$:

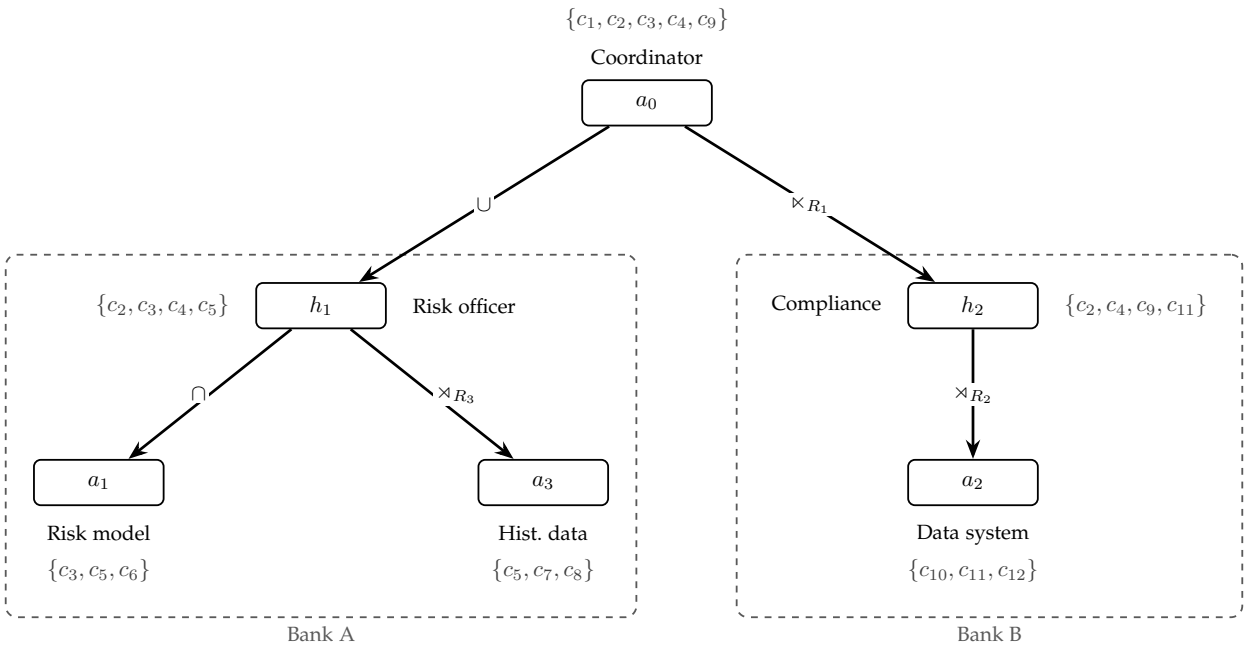
Credential	Meaning
c_1	Stress test coordination authority
c_2	Regulatory reporting certification
c_3	Risk assessment methodology
c_4	AML clearance
c_5	Bank A internal systems access
c_6	VaR model certification
c_7	Backtesting authorization
c_8	Historical loss database access
c_9	Cross-organizational data sharing agreement
c_{10}	Bank B customer exposure data
c_{11}	Sanctions screening certification
c_{12}	Bank B internal audit access

Intrinsic credential sets:

$$\sigma(a_0) = \{c_1, c_2, c_3, c_4, c_9\} \quad \text{(coordination, reporting, methodology, AML, data sharing)} \tag{S40}$$
$$\sigma(h_1) = \{c_2, c_3, c_4, c_5\} \quad \text{(reporting, methodology, AML, Bank A access)} \tag{S41}$$
$$\sigma(a_1) = \{c_3, c_5, c_6\} \quad \text{(methodology, Bank A access, VaR)} \tag{S42}$$
$$\sigma(a_3) = \{c_5, c_7, c_8\} \quad \text{(Bank A access, backtesting, historical data)} \tag{S43}$$
$$\sigma(h_2) = \{c_2, c_4, c_9, c_{11}\} \quad \text{(reporting, AML, data sharing, sanctions)} \tag{S44}$$
$$\sigma(a_2) = \{c_{10}, c_{11}, c_{12}\} \quad \text{(exposure data, sanctions, audit access)} \tag{S45}$$

Graph: A rooted tree with a_0 at the root and two subtrees. The left subtree covers Bank A's risk function: $a_0 \to h_1$ via *union* (pooling risk expertise), $h_1 \to a_1$ via *intersection* (mutual validation of risk methodology), $h_1 \to a_3$ via *right join* with $R_3 = \{c_3, c_5\}$ (the data repository's credentials are primary; the risk officer contributes only methodology and systems access). The right subtree covers Bank B's compliance function: $a_0 \to h_2$ via *left join* with $R_1 = \{c_2, c_4, c_{11}\}$ (the coordinator's credentials are primary; Bank B's compliance officer contributes only regulatory credentials), $h_2 \to a_2$ via *right join* with $R_2 = \{c_9\}$ (the data system's credentials are primary; the compliance officer contributes only the data sharing agreement).

Dashed boxes indicate organizational boundaries, as in Example 3. This is the first example using all four fundamental operators on a single graph, with depth-two composition on both branches.



S9.0.2 Credential Propagation

$$\hat{\sigma}(a_0) = \sigma(a_0) = \{c_1, c_2, c_3, c_4, c_9\} \quad \text{(5 credentials)} \tag{S46}$$

$$\begin{aligned}\hat{\sigma}(h_1) &= \hat{\sigma}(a_0) \cup \sigma(h_1) \\ &= \{c_1, c_2, c_3, c_4, c_9\} \cup \{c_2, c_3, c_4, c_5\} \\ &= \{c_1, c_2, c_3, c_4, c_5, c_9\} \quad \text{(6 credentials)}\end{aligned} \tag{S47}$$

$$\begin{aligned}\hat{\sigma}(a_1) &= \hat{\sigma}(h_1) \cap \sigma(a_1) \\ &= \{c_1, c_2, c_3, c_4, c_5, c_9\} \cap \{c_3, c_5, c_6\} \\ &= \{c_3, c_5\} \quad \text{(2 credentials)}\end{aligned} \tag{S48}$$

$$\begin{aligned}\hat{\sigma}(a_3) &= \sigma(a_3) \cup (\hat{\sigma}(h_1) \cap R_3) \\ &= \{c_5, c_7, c_8\} \cup (\{c_1, c_2, c_3, c_4, c_5, c_9\} \cap \{c_3, c_5\}) \\ &= \{c_5, c_7, c_8\} \cup \{c_3, c_5\} = \{c_3, c_5, c_7, c_8\} \quad \text{(4 credentials)}\end{aligned} \tag{S49}$$

$$\begin{aligned}\hat{\sigma}(h_2) &= \hat{\sigma}(a_0) \cup (\sigma(h_2) \cap R_1) \\ &= \{c_1, c_2, c_3, c_4, c_9\} \cup (\{c_2, c_4, c_9, c_{11}\} \cap \{c_2, c_4, c_{11}\}) \\ &= \{c_1, c_2, c_3, c_4, c_9\} \cup \{c_2, c_4, c_{11}\} \\ &= \{c_1, c_2, c_3, c_4, c_9, c_{11}\} \quad \text{(6 credentials)}\end{aligned} \tag{S50}$$

$$\begin{aligned}\hat{\sigma}(a_2) &= \sigma(a_2) \cup (\hat{\sigma}(h_2) \cap R_2) \\ &= \{c_{10}, c_{11}, c_{12}\} \cup (\{c_1, c_2, c_3, c_4, c_9, c_{11}\} \cap \{c_9\}) \\ &= \{c_{10}, c_{11}, c_{12}\} \cup \{c_9\} = \{c_9, c_{10}, c_{11}, c_{12}\} \quad \text{(4 credentials)}\end{aligned} \tag{S51}$$

Remark S38 (Depth-two composition)**.** *The effective credential sets at the leaf nodes reflect two stages of operator composition. The risk model a_1 receives credentials that were first* expanded *by union (at $a_0 \to h_1$) and then* contracted *by intersection (at $h_1 \to a_1$): union gives h_1 six credentials, but intersection with a_1's three credentials leaves only two. The VaR model certification c_6—the risk model's unique credential—is* destroyed *by the intersection because h_1 does not hold it. By contrast, the historical data repository a_3 receives credentials through a right join that* preserves *its unique data credentials (c_7, c_8) while adding the risk officer's methodology credential c_3 through the filter R_3. Two siblings of the same parent, connected by different operators, with dramatically different effective credential sets: 2 versus 4.*

S9.0.3 Coalition Game and Shapley Analysis

We use the shared-outcome value function introduced in Example 3: $W(\hat{\boldsymbol{\sigma}}) = |\bigcup_{a \in S} \hat{\sigma}(a)|$—the stress test's value depends on the total credential coverage across all participants.

Selected coalition values: With six actors, the characteristic function has $2^6 - 1 = 63$ non-empty coalitions. We report a representative selection; the complete characteristic function and Shapley derivation are given in Appendix S10.

Coalition S	$v(S)$	Note
$\{a_0\}$	5	Coordinator alone
$\{h_1\}$	4	Risk officer alone
$\{a_0, h_1\}$	6	Union adds only c_5 (heavy overlap)
$\{a_0, a_1\}$	7	Disconnected; $5 + 3 = 8$? No: 7 (shared c_3)
$\{a_1, h_1\}$	4	Intersection destroys a_1's c_6
$\{a_0, h_1, a_1\}$	6	Adding h_1 to $\{a_0, a_1\}$: value drops by 1
$\{a_1, a_3\}$	5	Disconnected siblings: $3 + 3 = 6$? No: 5 (shared c_5)
$\{a_0, a_2, a_3\}$	11	Three AI agents, no humans: highest triple
$N \setminus \{h_1\}$	12	**Grand coalition minus risk officer**
N (grand coalition)	11	Full six-actor coalition

Remark S39 (The grand coalition is not optimal)**.** *The most striking result: $v(N \setminus \{h_1\}) = 12 > 11 = v(N)$. Removing the human risk officer from the grand coalition* increases *its value. The mechanism is the same as Example 2 but amplified by depth-two composition: when h_1 is present, the intersection edge to a_1 activates, destroying c_6 (VaR model certification). When h_1 is absent, a_0 and a_1 are disconnected, and a_1 contributes its full standalone set $\{c_3, c_5, c_6\}$ independently. The human manager's role as an intermediary is the problem—the union edge to a_0 contributes little (heavy credential overlap), while the intersection edge to a_1 destroys a unique credential.*

As in Example 2, this result concerns the credential propagation graph*, not the operational workflow. A regulatory stress test may* require *a human risk officer regardless of the algebraic effect; the same holds of the all-agent coalition $\{a_0, a_1, a_2, a_3\}$, whose value 12 is a diagnosis, not an endorsement: it says that the four agents between them hold every credential, so that the two humans contribute no credential and only their filters change the outcome. Whether a coalition is* permitted *to act is a constraint outside the valuation: a mandated overseer or attestation enters as a requirement Q on the coalition (Definition 6.2 of the main text), and $v(S)$ records what S can do, never that it may. The finding's practical implication is that the operator connecting h_1 to a_1 should be reconsidered: replacing intersection with a join that preserves h_1's oversight without filtering a_1's unique credentials would retain the human in the workflow while eliminating the algebraic penalty.*

Shapley values:

Actor	Role	$\lvert\sigma\rvert$	ϕ	Rank
a_0	AI coordinator (root)	5	2.50	1st (tied)
a_2	AI data system, Bank B	3	2.50	1st (tied)
a_3	AI historical data, Bank A	3	2.33	3rd
h_2	Human compliance, Bank B	4	1.67	4th
a_1	AI risk model, Bank A	3	1.17	5th
h_1	Human risk officer, Bank A	4	0.83	6th (last)

Verification: $2.50 + 2.50 + 2.33 + 1.67 + 1.17 + 0.83 = 11.00 = v(N)$ ✓.

Remark S40 (Five findings from the Shapley analysis)**.**

(i) The human risk officer is the least valuable actor. $\phi_{h_1} = 0.83$—*last among six actors, despite holding 4 credentials (more than any AI agent except a_0). The union edge to a_0 contributes little because their credential sets overlap heavily ($\{c_2, c_3, c_4\}$ are shared). The intersection edge to a_1 is actively destructive. The net effect: h_1's presence hurts more coalitions than it helps.*

(ii) The AI coordinator and Bank B data system are tied for first. $\phi_{a_0} = \phi_{a_2} = 2.50$*, despite being at opposite ends of the tree (root vs. deepest node), in different organizations, and connected through different operators. The coordinator contributes unique authority credentials (c_1); the data system contributes unique Bank B data (c_{10}, c_{12}). Both are irreplaceable.*

(iii) Operator determines sibling inequality. *a_1 and a_3 are siblings under the same parent h_1, both holding 3 credentials. Yet $\phi_{a_3} = 2.33$ while $\phi_{a_1} = 1.17$—a factor of two. The only difference is the operator: intersection destroys a_1's unique credential c_6, while the right join preserves a_3's unique credentials c_7 and c_8.*

(iv) The overseers are outvalued by the actors they oversee. *h_1 ranks last; h_2 ranks fourth. In both cases, the actors they manage (a_1/a_3 and a_2 respectively) collectively contribute more. This is not because the overseers lack credentials, and nothing in the computation refers to their being human: it is because the operators on their edges either create overlap (union) or impose filtering (left join). An AI actor in the same position with the same edges would be valued identically.*

(v) h_1 has negative marginal contribution to every coalition containing a_0 and a_1. *In 12 of the 16 coalitions where h_1's marginal can be computed, $\Delta_{h_1}(S) = -1$. Adding the risk officer is destructive whenever the intersection edge activates. This is the most systematic negative-marginal result in the paper: not an isolated case but a* pattern*.*

S10 Full Computation for Example 4

S10.1 Characteristic Function

The shared-outcome value function is $v(S) = |\bigcup_{a \in S} \hat{\sigma}(a)|$, where effective credential sets are computed using only edges internal to S and disconnected actors contribute their intrinsic credential sets. All 63 non-empty coalitions are listed below,

grouped by size and sorted by value within each group.

Singletons (6):

S	$v(S)$
$\{a_0\}$	5
$\{h_1\}$	4
$\{h_2\}$	4
$\{a_1\}$	3
$\{a_3\}$	3
$\{a_2\}$	3

Pairs (15):

S	$v(S)$	Active edges
$\{a_0, a_3\}$	8	(none: disconnected)
$\{a_0, a_2\}$	8	(none: disconnected)
$\{a_0, a_1\}$	7	(none: disconnected)
$\{h_1, a_2\}$	7	(none: disconnected)
$\{a_1, h_2\}$	7	(none: disconnected)
$\{a_3, h_2\}$	7	(none: disconnected)
$\{a_0, h_1\}$	6	$\cup$
$\{a_0, h_2\}$	6	$\ltimes_{R_1}$
$\{h_1, a_3\}$	6	$\rtimes_{R_3}$
$\{h_1, h_2\}$	6	(none: disconnected)
$\{a_1, a_2\}$	6	(none: disconnected)
$\{a_3, a_2\}$	6	(none: disconnected)
$\{h_2, a_2\}$	6	$\rtimes_{R_2}$
$\{a_1, a_3\}$	5	(none: disconnected siblings)
$\{h_1, a_1\}$	4	$\cap$

Note that $v(\{h_1, a_1\}) = 4$: union of effective sets is $\{c_2, c_3, c_4, c_5\}$ because the intersection edge destroys a_1's unique credential c_6, and the effective set at a_1 reduces to $\{c_3, c_5\}$, which is a subset of h_1's intrinsic credentials.

Triples (20):

S	$v(S)$
$\{a_0, a_3, a_2\}$	11
$\{a_0, a_1, a_2\}$	10
$\{a_0, h_1, a_2\}$	9
$\{a_0, a_1, a_3\}$	9
$\{a_0, a_3, h_2\}$	9
$\{h_1, a_3, a_2\}$	9
$\{a_1, a_3, h_2\}$	9
$\{a_1, h_2, a_2\}$	9
$\{a_3, h_2, a_2\}$	9
$\{a_0, h_1, a_3\}$	8
$\{a_0, a_1, h_2\}$	8
$\{a_0, h_2, a_2\}$	8
$\{h_1, a_3, h_2\}$	8
$\{h_1, h_2, a_2\}$	8
$\{a_1, a_3, a_2\}$	8
$\{a_0, h_1, h_2\}$	7
$\{h_1, a_1, a_2\}$	7
$\{a_0, h_1, a_1\}$	6
$\{h_1, a_1, a_3\}$	6
$\{h_1, a_1, h_2\}$	6

Quadruples (15):

S	$v(S)$
$\{a_0, a_1, a_3, a_2\}$	12
$\{a_0, h_1, a_3, a_2\}$	11
$\{a_0, a_3, h_2, a_2\}$	11
$\{a_1, a_3, h_2, a_2\}$	11
$\{a_0, a_1, a_3, h_2\}$	10
$\{a_0, a_1, h_2, a_2\}$	10
$\{h_1, a_3, h_2, a_2\}$	10
$\{a_0, h_1, a_1, a_2\}$	9
$\{a_0, h_1, a_3, h_2\}$	9
$\{a_0, h_1, h_2, a_2\}$	9
$\{h_1, a_1, a_3, a_2\}$	9
$\{a_0, h_1, a_1, a_3\}$	8
$\{h_1, a_1, a_3, h_2\}$	8
$\{h_1, a_1, h_2, a_2\}$	8
$\{a_0, h_1, a_1, h_2\}$	7

Note: $v(\{a_0, a_1, a_3, a_2\}) = 12$ (all 12 credentials covered). This four-AI coalition achieves full coverage *without either human*.

Quintuples (6):

S	$v(S)$
$N \setminus \{h_1\}$	12
$N \setminus \{h_2\}$	11
$N \setminus \{a_2\}$	9
$N \setminus \{a_3\}$	9
$N \setminus \{a_1\}$	11
$N \setminus \{a_0\}$	10

Key observation: $v(N \setminus \{h_1\}) = 12 > 11 = v(N)$. Removing the risk officer *increases* the coalition's value.

Grand coalition (1): $v(N) = v(\{a_0, h_1, a_1, a_3, h_2, a_2\}) = 11$.

The grand coalition does not achieve full credential coverage because the intersection edge $h_1 \to a_1$ destroys c_6 (VaR model certification). Without h_1, actors a_0 and a_1 are disconnected, a_1 contributes its full intrinsic set including c_6, and the coalition covers all 12 credentials.

S10.2 Shapley Value Computation

The Shapley value for each actor is computed using the standard formula with $n = 6$ actors ($6! = 720$ orderings):

$$\phi_i = \sum_{S \subseteq \mathcal{A} \setminus \{i\}} \frac{|S|!\,(5 - |S|)!}{720}\,\Delta_i(S)$$

where $\Delta_i(S) = v(S \cup \{i\}) - v(S)$.

Each actor's marginal contribution must be computed for all $2^5 = 32$ subsets of the remaining five actors. Rather than reproduce all 192 marginal contributions, we provide the weighted sums by subset size and the final values.

ϕ_{a_0}:: Weighted sum of marginals: $\frac{120\cdot5+24\cdot18+12\cdot25+12\cdot17+24\cdot6+120\cdot1}{720} = \frac{1800}{720} = \frac{5}{2}$.

ϕ_{h_1}:: Weighted sum of marginals: $\frac{120\cdot4+24\cdot11+12\cdot8+12\cdot(-2)+24\cdot(-4)+120\cdot(-1)}{720} = \frac{600}{720} = \frac{5}{6}$.

Note the negative contributions at sizes 3, 4, and 5: h_1's marginal is -1 for every coalition containing both a_0 and a_1 (12 coalitions total), confirming the systematic destructive pattern identified in Section 4.2 of the main text.

ϕ_{a_1}:: Weighted sum: $\frac{120\cdot3+24\cdot10+12\cdot12+12\cdot6+24\cdot1+120\cdot0}{720} = \frac{840}{720} = \frac{7}{6}$.

ϕ_{a_3}:: Weighted sum: $\frac{120\cdot3+24\cdot13+12\cdot23+12\cdot21+24\cdot10+120\cdot2}{720} = \frac{1680}{720} = \frac{7}{3}$.

ϕ_{h_2}:: Weighted sum: $\frac{120\cdot4+24\cdot14+12\cdot18+12\cdot10+24\cdot2+120\cdot0}{720} = \frac{1200}{720} = \frac{5}{3}$.

ϕ_{a_2}:: Weighted sum: $\frac{120\cdot3+24\cdot14+12\cdot26+12\cdot24+24\cdot11+120\cdot2}{720} = \frac{1800}{720} = \frac{5}{2}$.

Verification: $\phi_{a_0} + \phi_{h_1} + \phi_{a_1} + \phi_{a_3} + \phi_{h_2} + \phi_{a_2} = \frac{5}{2} + \frac{5}{6} + \frac{7}{6} + \frac{7}{3} + \frac{5}{3} + \frac{5}{2} = \frac{15+5+7+14+10+15}{6} = \frac{66}{6} = 11 = v(N)$. ✓

S11 Computational Evaluation

The results of the main text are theorems; the experiments in this section illustrate them on random instances and answer two questions a practitioner will ask first: how often does the phenomenon occur, and what do the audit and detection procedures cost at scale. All numbers are produced by the scripts `exp1_prevalence.py`, `exp1b_redundancy.py`, `exp2_scale.py`, `exp3_fpt.py`, `exp3b_fpt_growth.py`, and `exp4_shapley.py` in the accompanying code (https://github.com/yryabov/game-of-peers, archived at https://doi.org/10.5281/zenodo.23002446); `test_examples.py` reproduces every number of the worked examples (Examples 1–4) as unit tests, and the `verify_*.py` scripts check the structural statements of the paper against exhaustive enumeration (see "Exhaustive checks" below). Timings are single-threaded pure Python without optimization; they are meant as orders of magnitude.

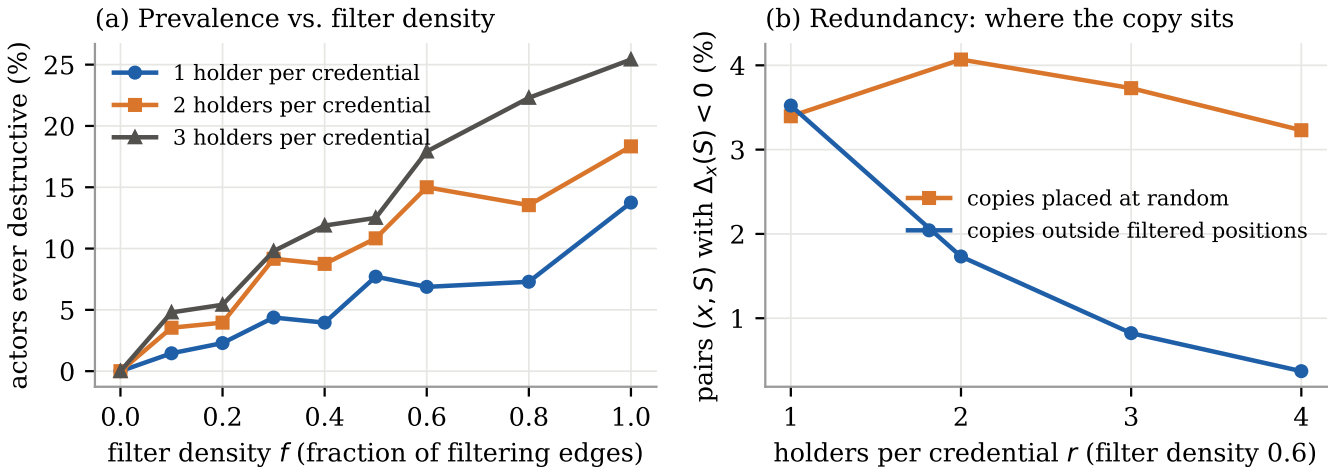


Fig. S1. Prevalence of destructive intermediaries on random trees. (a) Fraction of actors that are destructive in at least one coalition (exact, Prop. 6.3) against filter density, for one, two, and three holders per credential ($n = 30$, $|\mathcal{C}| = 6$). (b) Fraction of sampled pairs (x, S) with $\Delta_x(S) < 0$ against redundancy at filter density 0.6, with additional copies placed at random versus outside filtered positions.

Random instances: A rooted tree on n actors is generated by attaching each actor to a uniformly random earlier actor (expected depth $\Theta(\log n)$). Each edge carries a *filtering* operator ($\cap$, or $\ltimes_R$ with a uniformly random R) with probability f, the *filter density*, and a preserving operator ($\cup$ or $\rtimes_R$) otherwise. Each of $|\mathcal{C}|$ credentials is held by r distinct actors chosen uniformly at random; r is the *redundancy*. Coalition value is shared-outcome cardinality.

Prevalence of destructive intermediaries: Fig. S1(a) reports, for $n = 30$, $|\mathcal{C}| = 6$, and 60 trees per point, the fraction of actors that are destructive in *some* coalition, decided exactly by the pair-context dynamic program of Proposition 6.3 of the main text. With no filtering edge no actor is ever destructive, as Theorem 4.1 requires; the fraction then grows roughly linearly in f, reaching 14% at $f = 1$ for unique holdings and 25% for three holders per credential. The last figure is the surprise: naive redundancy makes matters *worse*. Fig. S1(b) separates the two mechanisms at $f = 0.6$ (80 trees, 40 random coalitions per actor). When the additional copies of each credential are placed at random, the rate of destructive pairs (x, S) stays flat near 3.5% while the fraction of ever-destructive actors climbs from 7% to 23%: every copy that lands at a filtered position is a new target for the filter, and a credential held by two filtered children of the same intermediary and by nobody else is destroyed together with both—exactly the case that the loss set L of Theorem 4.2 accounts for and the per-child form misses. When the copies are placed outside filtered positions, the rate of destructive pairs falls tenfold, from 3.5% to 0.4% at $r = 4$, while the fraction of ever-destructive actors stays between 5% and 8%. Both curves are what Proposition 4.1 of the main text predicts: a protected holder guards the credential in every coalition that contains it (part (i), the tenfold drop), but a single unprotected holder suffices for some coalition to lose it (part (ii), the flat ever-destructive rate, since the original holder is left where it was). The "protected" placement of the experiment uses the simplest protected position, an actor whose incoming edge is preserving. This is the precise content of the design principle "redundancy protects only from protected positions" (Section 7.1 of the main text).

Exhaustive checks: The `verify_*.py` scripts compare the stated results with exhaustive enumeration of all coalitions on small random instances. Theorem 4.2: on 3701 random pairs (x, S) satisfying its structural condition (trees with $n \leq 8$, all four operators), the formula, the identity $L = \hat{\sigma}_S \setminus \hat{\sigma}_T$, and the per-child form (529 applicable pairs) held without exception (`verify_criterion.py`). Proposition 5.3: pure equilibria existed in all 3000 random modular games over the four operators; Proposition 6.6: the prefix algorithm matched the exhaustive optimum on all 600 random mixed-operator trees (`verify_modular_mcact.py`); Proposition S6 held on 5690 random graphs, 2133 of them cyclic (`verify_agg_indep_cycles.py`). For the protection criterion (`verify_protection.py`, `verify_protection_general.py`): on 400 random trees ($n \leq 7$, $|\mathcal{C}| \leq 4$, all four operators) the three claims of Proposition 4.1 held in all 1216 credential-level checks; on 1500 random acyclic graphs the "every parent" form (Proposition S1(ii)) held in all 2951 checks; on 2000 random graphs with cycles sufficiency held in all 2665 checks while the converse failed in four cyclic instances, as Proposition S1(iii) says it must; and on 60 random formulas the necklace of Proposition 4.3 lost its credential exactly when the formula was satisfiable. The entry characterization (Lemma S2) was checked on 15,949 coalition–credential pairs.

Audit at scale: Fig. S2(a) times a single marginal contribution $\Delta_x(S)$—two coalition evaluations—on trees of 10^3 to 10^5 actors with $|\mathcal{C}| \in \{20, 50, 100\}$ and random coalitions of half the actors. The cost is linear in n and insensitive to $|\mathcal{C}|$ in this range (0.5 ms at $n = 10^3$, 0.18 s at $n = 10^5$), as Proposition 6.1 states; a workflow of 10^5 sessional tokens is audited exactly in well under a second per query.

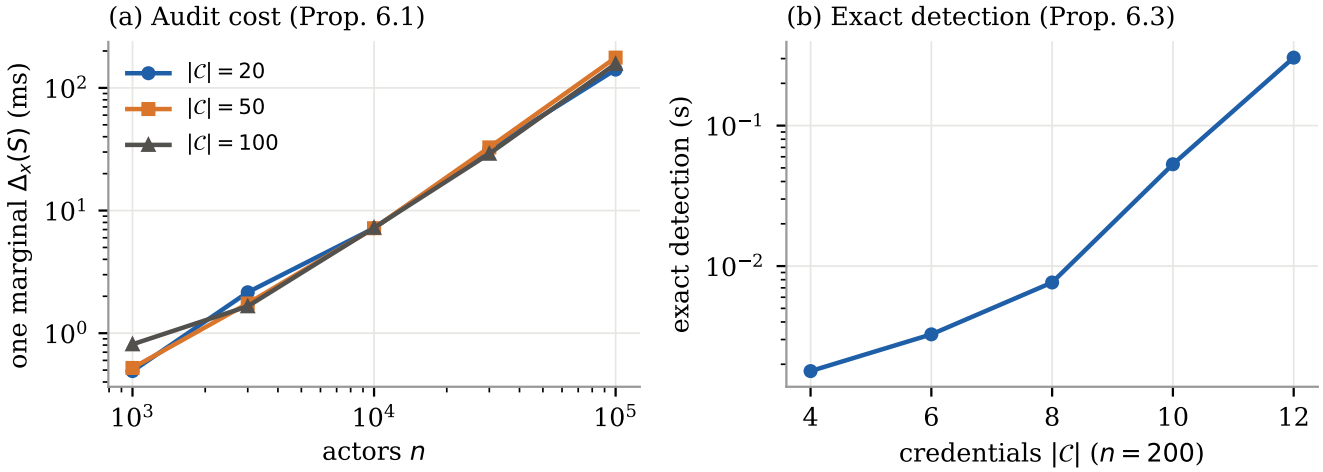


Fig. S2. Cost of audit and detection. (a) Time of one marginal contribution on random trees, log–log, three credential-universe sizes. (b) Time of exact destructiveness detection at $n = 200$ against $|\mathcal{C}|$, log scale.

Exact detection: Fig. S2(b) times the exact decision of DESTRUCTIVE-ACTOR by the dynamic program of Proposition 6.3 at $n = 200$ as $|\mathcal{C}|$ grows: 2 ms at $|\mathcal{C}| = 4$, 8 ms at 8, 0.3 s at 12, against 2^{199} coalitions for brute force. The growth is exponential in $|\mathcal{C}|$ as the bound $2^{O(|\mathcal{C}|)}$ predicts, and negligible for the credential universes of the worked examples ($|\mathcal{C}| \leq 12$). The program was cross-checked against exhaustive enumeration on 60 instances with $n \leq 14$.

Shapley sampling: On 20 random instances with $n = 30$ and $|\mathcal{C}| = 8$, the exact Shapley value of a credential holder (computed by the dynamic program with multiplicities, 0.12 s on average; mean $|\phi_x| = 0.76$) was compared with the sampling estimator of Proposition 6.4. The mean absolute error over ten runs was 0.039 with 10^2 permutations, 0.014 with 10^3, and 0.004 with 10^4; the largest error observed with 10^4 permutations was 0.018, comfortably inside the Hoeffding bound.

REFERENCES

[S1] A. Tarski, "A lattice-theoretical fixpoint theorem and its applications," *Pacific Journal of Mathematics*, vol. 5, no. 2, pp. 285–309, 1955.
[S2] D. M. Topkis, *Supermodularity and Complementarity*. Princeton, NJ: Princeton University Press, 1998.
[S3] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, 1996.
[S4] R. D. Carr, S. Doddi, G. Konjevod, and M. V. Marathe, "On the red-blue set cover problem," in *Proc. 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 345–353, 2000.
[S5] L. G. Valiant, "The complexity of enumeration and reliability problems," *SIAM Journal on Computing*, vol. 8, no. 3, pp. 410–421, 1979.
[S6] X. Deng and C. H. Papadimitriou, "On the complexity of cooperative solution concepts," *Mathematics of Operations Research*, vol. 19, no. 2, pp. 257–266, 1994.
[S7] J. Castro, D. Gómez, and J. Tejada, "Polynomial calculation of the Shapley value based on sampling," *Computers & Operations Research*, vol. 36, no. 5, pp. 1726–1730, 2009.
[S8] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations* (R. E. Miller and J. W. Thatcher, eds.), pp. 85–103, Plenum Press, 1972.
[S9] V. Chvátal, "A greedy heuristic for the set-covering problem," *Mathematics of Operations Research*, vol. 4, no. 3, pp. 233–235, 1979.
[S10] U. Feige, "A threshold of $\ln n$ for approximating set cover," *Journal of the ACM*, vol. 45, no. 4, pp. 634–652, 1998.
[S11] N. Li, J. C. Mitchell, and W. H. Winsborough, "Design of a role-based trust-management framework," in *Proc. 2002 IEEE Symposium on Security and Privacy*, pp. 114–130, 2002.
[S12] N. Li and J. C. Mitchell, "Datalog with constraints: A foundation for trust management languages," in *Proc. 5th Int. Symp. Practical Aspects of Declarative Languages (PADL)*, LNCS vol. 2562, pp. 58–73, 2003.
[S13] M. A. Heller, "The tragedy of the anticommons: Property in the transition from Marx to markets," *Harvard Law Review*, vol. 111, no. 3, pp. 621–688, 1998.
[S14] J. F. Nash, "Equilibrium points in n-person games," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 36, no. 1, pp. 48–49, 1950.
[S15] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*, Cambridge University Press, 2007.
[S16] "DAO-Agent: Zero knowledge-verified incentives for multi-agent coordination," *arXiv preprint arXiv:2512.20973*, 2025.
[S17] J. R. Green and J.-J. Laffont, "Characterization of satisfactory mechanisms for the revelation of preferences for public goods," *Econometrica*, vol. 45, no. 2, pp. 427–438, 1977.
[S18] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
[S19] I. Terrucha, E. Fernández Domingos, R. Suchon, F. C. Santos, P. Simoens, and T. Lenaerts, "Humans program artificial delegates to accurately solve collective-risk dilemmas but lack precision," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 122, no. 25, e2319942121, 2025.
[S20] C. M. de Melo, S. Marsella, and J. Gratch, "Human cooperation when acting through autonomous machines," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 8, pp. 3482–3487, 2019.
[S21] T. G. Griffin, F. B. Shepherd, and G. Wilfong, "The stable paths problem and interdomain routing," *IEEE/ACM Transactions on Networking*, vol. 10, no. 2, pp. 232–243, 2002.
[S22] "Authenticated workflows: A systems approach to protecting agentic AI," *arXiv preprint arXiv:2602.10465*, 2026.
[S23] H. Huang et al., "A novel zero-trust identity framework for agentic AI: Decentralized authentication and fine-grained access control," *arXiv preprint arXiv:2504.13538*, 2025.