A Resource-Aware Structural Causal Semantics for Feasibility Analysis of Counterfactuals

Pinaki Chakraborty[1], David Pym[1,2] and Tristan Caulfield[1]

[1]University College London, London, United Kingdom
[2]Institute of Philosophy, University of London, London, United Kingdom

Abstract

Structural causal models (SCMs) are the default substrate for counterfactual explanation, responsibility, and causal recourse in the Halpern-Pearl account of actual causation. In SCMs, interventions quantify over arbitrary assignments to endogenous variables. However, in socio-technical systems, agent actions (and hence responsibility) are constrained by procedures, authority, permissions, and coordination requirements. Thus counterfactual 'alternatives' may include states that the relevant agent could not have feasibly executed. We characterize this intervention–feasibility gap and show how it can yield misplaced responsibility attributions or infeasible recourse recommendations even when the underlying SCM is descriptively adequate. We propose a principled restriction of intervention admissibility via *Resource-Aware SCMs* (R-SCMs). Without abandoning SCM semantics, we equip models with agent-controlled endogenous variables, enablement predicates evaluated against the acting agent's observation, and consumable resources modelled by a partial commutative monoid. We interpret agent-relative counterfactual feasibility via executable, resource-consuming intervention traces and prove basic properties of R-SCMs (well-definedness, monotonicity, conservativity under trivial feasibility). Finally, we prove NP-completeness of the bounded-horizon feasibility decision problem for R-SCMs.

Keywords

Causality, Resource-Sensitive Reasoning, Feasibility, Responsibility

1. Introduction

Responsibility and recourse questions about modern AI- and software-mediated decision *workflows* increasingly arise in socio-technical systems that are *governance-oriented* in a concrete sense: decision-making is organized around explicit roles and interfaces, constrained by procedures and policies. These settings are often audited after the fact, and the question whether an agent 'could have done otherwise' is not merely a question about logical possibility; it is a question about what the agent was enabled, authorized, informed, and resourced to do. This perspective is familiar from bounded rationality [1, 2], organizational routines [3], and accounts of meaningful human control in safety-critical automation [4].

Pearl-style structural causal models (SCMs), together with the Halpern-Pearl (HP) account of actual causation and graded responsibility [5, 6, 7], provide the dominant formal bridge from a descriptive model to counterfactual explanation, responsibility, and recourse. Their intervention operator do(·) is intentionally liberal: for an endogenous variable X and value $x \in \mathcal{R}(X)$, do(X=x) is a well-defined semantic operation on the equations. HP contingencies inherit this freedom, since auxiliary variables may be fixed to chosen settings to reveal counterfactual dependence. In socio-technical workflows, however, this language ranges over both genuine control moves and assignments no relevant agent could execute. We call the resulting mismatch the *intervention–feasibility gap*. It can support responsibility claims or recourse recommendations that are formally valid but procedurally impossible, a concern also visible in work on actionable and intervention-aware recourse [8, 9, 10].

One response is to move to an action-theoretic representation, where preconditions, permissions, timing, and coordination requirements are part of the action signature. The KR literature offers many such accounts and bridges to causal reasoning [11, 12, 13, 14, 15, 16]. Our aim is different: we keep the SCM as the underlying causal substrate and make only the admissibility of counterfactual alternatives

24th International Workshop on Nonmonotonic Reasoning, July 17–19, 2026, Lisbon, Portugal
pinaki.chakraborty.22@ucl.ac.uk (P. Chakraborty); david.pym@sas.ac.uk (D. Pym); t.caulfield@ucl.ac.uk (T. Caulfield)

agent-relative. An agent 'could have done otherwise' only when there exists an executable trace of enabled primitive interventions under that agent's observation and resources.

The present contribution is best read as a causality-and-action account of admissible counterfactuals in the Halpern-Pearl paradigm. We introduce Resource-Aware SCMs (R-SCMs), a conservative layer over SCM semantics for distinguishing formal interventions from alternatives that an agent could actually realise under procedural and resource constraints. In responsibility attribution and explanation, this matters because an explanation or recommendation should correspond to an executable procedure: the means of intervention may be limited, exclusive, or consumed by the attempt itself. A counterfactual is therefore available to an agent or coalition only when the corresponding controlled interventions form an enabled, resource-paying trace whose accumulated effect makes the target formula true. A partial composition operator, together with its induced preorder, is used to represent incompatibility and minimality.

In Section 2, we situate the approach relative to action-theoretic causation, agency and responsibility logics, resource-bounded ability, and causal recourse. Section 3 gives minimal examples of the intervention–feasibility gap in responsibility and recourse. Section 4 defines R-SCMs, feasible traces, coalitional feasibility, and responsibility witnesses. Section 5 proves well-definedness, monotonicity, conservativity under trivial feasibility, and NP-completeness of bounded-horizon feasible achievement and prevention. Section 6 concludes with limitations and proof-certificate directions for auditable feasibility witnesses.

2. Related Work

Our starting point is Pearl's structural-equations account of causality [5, 17] and the HP framework for actual causation [6]. The agent-relative stakes are closest to work on responsibility and blame [7, 18], but our target is not a refinement of HP causation itself. We ask which interventions should be admissible for an agent when the model is used for responsibility or recourse. Constraint and normality approaches also restrict counterfactual reasoning. Causal models with constraints rule out impossible states globally [19], while normality orderings rank candidate contingencies by typicality [20]. R-SCMs instead require an agent-relative, resource-feasible execution witness.

Action-theoretic accounts make feasibility explicit by construction. Work connecting SCMs to action languages, independent choice logic, and situation-calculus-style accounts of causation shows that structural and action-based representations can be related [11, 14, 12, 13]. Related approaches formalize explanation and achievement causation in action languages and action theory, including ASP-based implementations [21, 16, 22, 15] and many other recent works which we omit for the sake of brevity. Bertossi's ASP-based work on counterfactual explanations for classification similarly allows domain knowledge to restrict the counterfactual space, including actionable refinements [23]. Unlike our framework, however, it does not make feasibility a first-class SCM-semantic notion grounded in agent-relative executable traces, observation, and consumable resources. These frameworks motivate our treatment of executable alternatives, but they typically begin with an action description whose preconditions already encode feasibility. We instead start with an SCM and add a restriction principle for admissible counterfactual antecedents.

Logical theories of responsibility in multi-agent systems commonly take agency primitives such as choices, strategies, game forms, or STIT-style abilities as basic [24, 25, 26]. Resource-bounded strategic logics and imperfect-information variants likewise internalize limits on what agents or coalitions can achieve [27, 28, 29]. R-SCMs provide a complementary SCM-based substrate: the witness for 'could have done otherwise' is an enabled intervention trace with explicit resource consumption. This also matches the motivation in causal recourse, where arbitrary feature-setting counterfactuals have been criticized in favor of actionable, intervention-aware recommendations [8, 9, 10]. In our setting, recourse is evaluated only over traces feasible under the subject's operative constraints.

3. Failure Modes of Unconstrained Counterfactual Reasoning

In this section, we discuss a couple of concrete issues with SCM-based analyses of responsibility and recourse. Because do(·) ranges over arbitrary endogenous variable assignments rather than an agent's executable options, these can certify those interventions which are infeasible in the actual situation. Given an SCM $\mathcal{M}$ and context $\vec{u}$, $(\mathcal{M}, \vec{u}) \models \varphi$ denotes truth in the unique solution, and $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(X{=}x)]\,\varphi$ denotes truth in the intervened model. Whether such an intervention corresponds to something an agent could actually do in the given world is *not* represented in $\mathcal{M}$ by default. We call an intervention $\mathrm{do}(X{=}x)$ *outcome-flipping* for an event B if $(\mathcal{M}, \vec{u}) \models B$ but $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(X{=}x)]\,\neg B$. The intervention–feasibility gap arises exactly when an outcome-flipping counterfactual is treated as a legitimate alternative in a responsibility or recourse argument despite being infeasible for the agent. As mentioned previously, our core theoretical development (Section 4) resolves this by restricting agent-relative counterfactual reasoning to those witnessed by enabled, resource-consuming execution traces.

Let A mean an alarm is shown, T that the response window is long enough, H that the human aborts, and B that a bad outcome occurs: $H := A \wedge T$ and $B := \neg H$. Assume $(\mathcal{M}, \vec{u}) \models A{=}1$ and $(\mathcal{M}, \vec{u}) \models T{=}0$, hence $H{=}0$ and $B{=}1$ (with $H{=}0$ indicating falsity and $B{=}1$ indicating that the bad outcome occurs). Unconstrained counterfactual reasoning permits $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(H{=}1)]\,\neg B$, inviting the reading that the human *could* have prevented the outcome. But when $T{=}0$, setting $H{=}1$ does not correspond to any executable option for the human in that situation: the intervention bypasses the very constraint captured by the model. A feasibility-aware analysis therefore shifts attention to what fixed T (automation speed, interface latency, early-warning design) rather than attributing responsibility to a non-existent option.

In causal recourse, a denied subject is given changes that would flip a decision [10, 30, 31]. Let $D := \mathbb{I}[\mathrm{score}(X) \geq \tau]$ be a deterministic decision rule over features X. A standard objective searches for a small feature move ΔX such that $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(X := X + \Delta X)]\,(D{=}1)$. An *outcome-flipping* do(·) antecedent is not a feasible option for the relevant agent, yet is treated as a legitimate recommendation. This is the intervention–feasibility gap in recourse form: the counterfactual is well-defined, indeed outcome-flipping for D, yet it does not correspond to any executable plan for the relevant agent. These examples motivate the conservative move of the next section. To reiterate, we keep the SCM's underlying structure intact, but restrict agent-relative counterfactual reasoning to those witnessed by *enabled procedures*.

4. A Resource-Aware Semantics for Feasible Counterfactual Reasoning

We take a procedural view of counterfactual alternatives: an agent 'could have done otherwise' only if there exists a concrete sequence of enabled actions that the agent could execute in the given situation, echoing Simon's emphasis on bounded rationality [1]. The underlying SCM semantics is unchanged: only the *admissible* counterfactual alternatives for agent-relative queries are restricted.

4.1. SCMs and Interventions

We assume deterministic acyclic structural causal models (SCMs). An SCM is a tuple $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{R}, F)$, where $\mathcal{U}$ are exogenous variables (contexts), $\mathcal{V}$ endogenous variables, $\mathcal{R}$ gives finite ranges, and $F = \{f_V\}_{V \in \mathcal{V}}$ assigns each $V \in \mathcal{V}$ a structural equation $V := f_V(\mathrm{Pa}(V), \vec{U})$. Here $\mathrm{Pa}(V) \subseteq \mathcal{V} \setminus \{V\}$ denotes the (endogenous) parents of V, i.e., the set of variables on which f_V depends. Equivalently, $\mathrm{Pa}(V)$ are the incoming neighbours of V in the acyclic directed causal graph induced by F. A *context* is an assignment $\vec{u}$ to $\mathcal{U}$. We write $\vec{U}$ for the exogenous-variable tuple and $\vec{u}$ for a particular valuation. For each context $\vec{u}$, acyclicity implies a unique endogenous valuation, written $\mathrm{Sol}_{\mathcal{M}}(\vec{u})$.

An *intervention map* is a finite partial function $I : \mathcal{V} \rightharpoonup \bigcup_{X \in \mathcal{V}} \mathcal{R}(X)$, with $I(X) \in \mathcal{R}(X)$, when defined. Let $\mathcal{M}_I$ be the intervened SCM obtained by replacing, for each $X \in \mathrm{dom}(I)$, the structural equation for X by the constant equation $X := I(X)$, leaving all other structural equations unchanged. We write $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(I)]\,\varphi$ to denote that φ is satisfied under the valuation $\mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$. We fix a finite

set of agents $\mathcal{A} = \{1, \dots, n\}$. The remaining definitions in this subsection add an operational layer to this standard SCM notation. Resource frames describe what can be spent or combined, cumulative intervention maps and states record what has already been done, primitive actions and traces describe executable procedures, and the final modality connects such procedures back to ordinary truth in intervened SCMs.

Definition 1 (Resource frame and residual)**.** *A* resource frame *is a partial commutative monoid* $\mathfrak{R} = (\mathcal{T}, \otimes, e)$*, where* $\mathcal{T}$ *is a set of resources,* $\otimes$ *is a partial binary operation on* $\mathcal{T}$*, and* $e \in \mathcal{T}$ *is a unit. We write* $t \otimes t' \downarrow$ *when the composition is defined and* $t \otimes t' \uparrow$ *otherwise. We require the following partial-monoid laws:*

1. *For every* $t \in \mathcal{T}$*,* $e \otimes t \downarrow$*,* $t \otimes e \downarrow$*, and* $e \otimes t = t \otimes e = t$*.*
2. *For all* $s, t \in \mathcal{T}$*,* $s \otimes t \downarrow$ *iff* $t \otimes s \downarrow$*, and, when defined,* $s \otimes t = t \otimes s$*.*
3. *For all* $r, s, t \in \mathcal{T}$*,* $(r \otimes s) \otimes t \downarrow$ *iff* $r \otimes (s \otimes t) \downarrow$*, and, when defined,* $(r \otimes s) \otimes t = r \otimes (s \otimes t)$*.*

Undefinedness represents a resource clash*, and definedness represents compatibility. The induced availability preorder is defined by* $c \preceq t$ *iff there exists* $r \in \mathcal{T}$ *such that* $c \otimes r \downarrow$ *and* $c \otimes r = t$*. When* $c \preceq t$*, any such* r *is called a* residual *of paying* c *from* t*. If* $c \preceq t$*, we also write* $t \succeq c$*.* □

Intuitively, an element $t \in \mathcal{T}$ can be viewed as a *set of tokens* representing the resources currently available, both *consumables* (e.g., time, budget, effort) that can be spent, and *exclusives* (e.g., locks, quorum shares, mutex) that cannot be jointly held or used in conflicting ways. This is the standard algebraic abstraction used to model consumable and exclusive resources: incompatibility is represented by partiality. For each agent $i \in \mathcal{A}$, we fix a set of *controlled variables* $C_i \subseteq \mathcal{V}$, representing endogenous variables that are, in principle, directly actuated by i. When I is an intervention map with $\mathrm{dom}(I) \subseteq C_i$, we call it an *$i$-control intervention.*

A trace is interpreted as a sequence of persistent (set-once discipline) overwrites of *controlled variables*, and after each trace prefix we evaluate the unique solution of the resulting intervened SCM. Feasibility is determined by *gates*: state-dependent predicates that decide whether a primitive action is available given the current context, resources, and constraints [32]. To avoid an omniscience assumption, gate predicates may depend on what the acting agent can observe at the current valuation, rather than on the full valuation itself. This reflects a standard epistemic concern in causal reasoning: causal and responsibility judgments are often conditioned on the information available to the agent [33].

Concretely, each agent i is equipped with an observation map $\mathsf{obs}_i : \prod_{V \in \mathcal{V}} \mathcal{R}(V) \to \mathcal{O}_i$, where $\mathcal{O}_i$ is the set of observations available to agent i; when the current cumulative intervention induces valuation $\vec{v}$, the enablement predicate (gate) for an i-action (Definition 4) is evaluated on observation $o_i = \mathsf{obs}_i(\vec{v})$. In what follows, we omit the subscript (or superscript) whenever it is convenient to do so. This ensures that feasibility claims used for responsibility attribution are judged against the information available to the agent. Organizational rules, norms, and contracts are treated uniformly as additional gate constraints.

The cumulative intervention map is the bookkeeping object for procedural counterfactuals. It records the persistent assignments made so far, while the set-once condition keeps a trace from rewriting the same controlled variable multiple times.

Definition 2 (Cumulative intervention map)**.** *Let us fix an initial intervention map* I *and define* $I_0 := I$*. For each* $t \in \{1, \dots, m\}$ *set* $I_t := I_{t-1} \cup \{X_t \mapsto x_t\}$*. Under the* set-once *discipline, we have* $X_t \notin \mathrm{dom}(I_{t-1})$*, so the union is well-defined. We call* I_t *the* cumulative intervention map *after* t *steps.* □

A state packages exactly the information needed to decide the next step of a procedure: the fixed exogenous context, the agents' remaining resources, and the intervention map accumulated so far. The associated valuation $\vec{v}_I$ is recomputed from the SCM after the accumulated intervention, so feasibility decisions can depend on the current causal consequences of earlier actions.

Definition 3 (State). *A* state *is a triple* $s = (\vec{u}, \rho, I)$ *where* $\vec{u}$ *is the exogenous context,* $\rho : \mathcal{A} \to \mathcal{T}$ *is a resource store, and* $I : \bigcup_{i \in \mathcal{A}} \mathcal{C}_i \rightharpoonup \bigcup_X \mathcal{R}(X)$ *is a finite partial map recording the current* cumulative intervention *on controlled variables. Given* I*, let* $\mathcal{M}_I$ *be the intervened SCM obtained by replacing, for each* $X \in \mathrm{dom}(I)$*, the structural equation for* X *by the constant equation* $X := I(X)$*, leaving all other structural equations unchanged. We write* $\vec{v}_I := \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ *for the induced endogenous valuation (which is well-defined for acyclic SCMs).* □

Primitive actions are the atomic moves from which executable counterfactual alternatives are built. They combine control authority ($X \in \mathcal{C}_i$), a proposed assignment $X := x$, a resource requirement, and a gate predicate that captures context-specific feasibility.

Definition 4 (Primitive action). *A* primitive action *is a pair* $\alpha^i = (i, X := x)$ *with* $i \in \mathcal{A}$*,* $X \in \mathcal{C}_i$*, and* $x \in \mathcal{R}(X)$*. In a given R-SCM, each action* $\alpha^i \in \mathrm{Act}$ *has a resource-token requirement* $\mathsf{cost}(\alpha) \in \mathcal{T}$ *and a gate predicate* $G_{\alpha^i}(\vec{u}, o_i, \rho, I) \in \{0, 1\}$*, evaluated at context* $\vec{u}$*, the acting agent's observation* o_i*, the full resource store* $\rho : \mathcal{A} \to \mathcal{T}$*, and the current cumulative intervention map* I*. Intuitively,* $G_{\alpha^i} = 1$ *entails both resource availability and any additional feasibility constraints. We require that if* $G_{\alpha^i}(\vec{u}, o_i, \rho, I) = 1$*, then* $\mathsf{cost}(\alpha) \preceq \rho(i)$*. A primitive action* $\alpha^i = (i, X := x)$ *is* enabled *at* $s = (\vec{u}, \rho, I)$*, written* $s \models \mathsf{en}(\alpha)$*, iff* $\alpha^i \in \mathrm{Act}$ *and, letting* $\vec{v}_I := \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ *and* $o_i := \mathsf{obs}_i(\vec{v}_I)$*, we have* $G_{\alpha^i}(\vec{u}, o_i, \rho, I) = 1$ *and* $X \notin \mathrm{dom}(I)$*.* □

The one-step relation gives the operational meaning of an enabled primitive action. Executing an action both pays its resource cost and extends the cumulative intervention map with the new assignment.

Definition 5 (One-step execution). *Let* $s = (\vec{u}, \rho, I)$ *and let* $\alpha = (j, X := x)$ *for some* $j \in \mathcal{A}$*. Let us write* $\vec{v} := \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ *and* $o := \mathsf{obs}_j(\vec{v})$*. We write* $s \xrightarrow{\alpha} s'$ *iff* $s' = (\vec{u}, \rho', I')$ *and:*

1. $\alpha \in \mathrm{Act}$ *and* $X \notin \mathrm{dom}(I)$*;*
2. $G_\alpha(\vec{u}, o, \rho, I) = 1$*;*
3. *there exists* $r \in \mathcal{T}$ *such that* $\mathsf{cost}(\alpha) \otimes r \downarrow$ *and* $\mathsf{cost}(\alpha) \otimes r = \rho(j)$*, with* $\rho'(j) = r$ *and* $\rho'(\ell) = \rho(\ell)$ *for all* $\ell \neq j$*;*
4. $I' = I \cup \{X \mapsto x\}$. □

Traces lift one-step execution from single actions to finite procedures. They are the witnesses used later for claims that an agent could feasibly have achieved or prevented an event.

Definition 6 (Trace). *A* trace *is a finite sequence of actions from* Act*. Operationally, executing a trace means applying the one-step transition relation (Definition 5) to its actions in order; after each prefix* $\alpha_1 ; \cdots ; \alpha_t$*, the accumulated assignments are recorded in the corresponding cumulative intervention map (Definition 2). A trace* $\sigma = \alpha_1 ; \cdots ; \alpha_m$ *is* executable *from a state* s *if there exist states* $s_0, \ldots, s_m$ *with* $s_0 = s$ *such that* $s_{t-1} \xrightarrow{\alpha_t} s_t$ *for each* $t = 1, \ldots, m$*. We write* $s \xrightarrow{\sigma} s_m$*, and if* $s_m = (\vec{u}, \rho_m, I_m)$ *we write* $I_\sigma := I_m$*.* □

An R-SCM collects the underlying causal model and the feasibility layer into a single structure. The structural equations still determine the consequences of interventions, while the action costs and gates determine which interventions can actually be reached by an agent.

Definition 7 (Resource-Aware SCM). *A* resource-aware SCM *(R-SCM) is a tuple*

$$\widehat{\mathcal{M}} = (\mathcal{M}, \mathcal{A}, \mathfrak{R}, (\mathcal{C}_i)_{i \in \mathcal{A}}, \mathrm{Act}, (\mathsf{obs}_i)_{i \in \mathcal{A}}, \mathsf{cost}, (G_\alpha)_{\alpha \in \mathrm{Act}})$$

where $\mathcal{M}$ *is a deterministic acyclic SCM,* $\mathcal{A}$ *is a finite agent set,* $\mathfrak{R} = (\mathcal{T}, \otimes, e)$ *is a resource frame, and* $\mathcal{C}_i \subseteq \mathcal{V}$ *is the set of endogenous variables controlled by agent* i*. The set* Act *consists of primitive actions compatible with these control sets, so that* $\mathrm{Act} \subseteq \{(i, X := x) \mid i \in \mathcal{A},\ X \in \mathcal{C}_i,\ x \in \mathcal{R}(X)\}$*. For each agent* i*, the observation map* $\mathsf{obs}_i : \prod_{V \in \mathcal{V}} \mathcal{R}(V) \to \mathcal{O}_i$ *specifies the information available to* i *at a valuation.*

The function $\mathsf{cost} : \mathrm{Act} \to \mathcal{T}$ *assigns resource-token requirements to primitive actions. For each action* $\alpha = (i, X := x) \in \mathrm{Act}$*, the gate predicate is* $G_\alpha : \mathrm{Ctx}(\mathcal{U}) \times \mathcal{O}_i \times (\mathcal{A} \to \mathcal{T}) \times \mathsf{Int} \to \{0, 1\}$ *where* $\mathrm{Ctx}(\mathcal{U})$ *is the set of exogenous contexts and* Int *is the set of finite cumulative intervention maps. Thus* $G_\alpha(\vec{u}, o, \rho, I)$ *determines whether* α *is admissible at context* $\vec{u}$*, acting-agent observation* $o \in \mathcal{O}_i$*, resource store* $\rho : \mathcal{A} \to \mathcal{T}$*, and cumulative intervention map* I*.* □

The feasible-counterfactual modality says that φ is feasible for agent i precisely when there is an executable i-only trace whose accumulated intervention makes φ true in the ordinary intervened SCM.

Definition 8 (Feasible-counterfactual modality)**.** *Let* $s = (\vec{u}, \rho, I)$ *be a state and let* φ *be an event formula over endogenous variables. We define* $(\widehat{\mathcal{M}}, s) \models \langle i \rangle \varphi$ *iff there exists an executable trace* σ *from* s*, consisting only of actions of agent* i*, such that for some final state* $s \xrightarrow{\sigma} (\vec{u}, \rho', I_\sigma)$ *we have* $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$*.* □

Following Halpern [6], we adopt the standard convention that an event E can be described by a propositional formula φ_E. We read $\langle i \rangle \neg \varphi_B$ as a feasible prevention of event B by i, and $\langle i \rangle \psi_D$ as feasible achievement of a desired outcome D by i. The responsibility notion used below refines feasible prevention by adding actual occurrence and resource-minimality conditions. Trivial feasibility marks the special case in which the resource-aware layer imposes no real restriction on an agent's primitive interventions. It is used below to show this semantics conservatively recovers ordinary interventions when every controlled primitive action is freely executable.

Definition 9 (Trivial feasibility)**.** *We say that feasibility is* trivial for agent i *iff, for every* $X \in \mathcal{C}_i$ *and every* $x \in \mathcal{R}(X)$*, the primitive action* $(i, X := x)$ *belongs to* Act*, has cost* e*, and for every state* $s = (\vec{u}, \rho, I)$ *with* $X \notin \mathrm{dom}(I)$*, letting* $\vec{v} := \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ *and* $o := \mathsf{obs}_i(\vec{v})$*, we have* $G_{(i, X := x)}(\vec{u}, o, \rho, I) = 1$*.* □

4.2. Coalitional Feasibility

Many socio-technical alternatives are available only through coordination. Since our primitive notion of feasibility is procedural (in the sense of witnessed by an executable trace), the minimal multi-agent extension is to allow traces whose steps are taken by different agents.

Definition 10 (Coalitional feasibility)**.** *Let* $C \subseteq \mathcal{A}$ *be a coalition. A* C-trace *is a trace* $\sigma = \alpha_1; \cdots; \alpha_m$ *such that each action* $\alpha_t = (i_t, X_t := x_t) \in \mathrm{Act}$ *has* $i_t \in C$*. A* C*-trace* σ *is* executable from *a state* $s = (\vec{u}, \rho, I)$ *iff there exist states* $s_0, \ldots, s_m$ *with* $s_0 = s$ *such that* $s_{t-1} \xrightarrow{\alpha_t} s_t$ *for each* $t = 1, \ldots, m$ *(Definition 5). If* $s_m = (\vec{u}, \rho_m, I_m)$*, we write* $I_\sigma := I_m$*.*

For an event formula φ *over endogenous variables, define the coalitional feasible-counterfactual modality by* $(\widehat{\mathcal{M}}, s) \models \langle C \rangle \varphi$ *iff there exists a* C*-trace* σ *executable from* s *such that, for some final state* $s \xrightarrow{\sigma} (\vec{u}, \rho', I_\sigma)$*, we have* $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$*. We write* $\langle i \rangle \varphi$ *as shorthand for* $\langle \{i\} \rangle \varphi$*.* □

Here coalitional ability is obtained purely by allowing different agents to take steps in the same executable procedure. Nothing else changes: resources remain per-agent components of ρ, feasibility remains encoded in the gates G_α, and gates are still evaluated on the acting agent's observation $o = \mathsf{obs}_{i_t}(\vec{v})$ at each step.

4.3. Responsibility

Definition 11. *Fix an R-SCM* $\widehat{\mathcal{M}}$*, an initial state* $s_0 = (\vec{u}, \rho, \emptyset)$*, an agent* i*, and an event formula* φ_B *describing event* B*. For an executable* i*-trace, define* $\mathsf{Cost}(\epsilon) := e$ *and, for* $\tau = \alpha_1; \cdots; \alpha_m$*,* $\mathsf{Cost}(\tau) := \mathsf{cost}(\alpha_1) \otimes \cdots \otimes \mathsf{cost}(\alpha_m)$*. For executable single-agent traces, this product is defined. This follows by induction from the residual equations in the one-step rule and the partial-monoid associativity law. A trace* $\sigma = \alpha_1; \cdots; \alpha_m$ *is a* responsibility witness for preventing B by i at s_0 *iff:*

1. *(*Potency*)* $(\mathcal{M}, \vec{u}) \models \varphi_B$*.*
2. *(*Feasible prevention*)* σ *is executable from* s_0*, each* α_t *has the form* $(i, X_t := x_t)$*, and for some final state* $s_0 \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$*, we have* $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \neg \varphi_B$*.*

3. *(*Token-minimality*) There is no executable i-trace σ' from s_0 such that $(\mathcal{M}_{I_{\sigma'}}, \vec{u}) \models \neg\varphi_B$ and $\mathsf{Cost}(\sigma') \prec \mathsf{Cost}(\sigma)$, where $\prec$ is the strict part of the preorder induced by $\otimes$.*

If no such σ exists, then we say that B is not preventable by i at s_0. □

Responsibility witnesses are evaluated from initial states. Non-empty intervention maps arise only while executing the counterfactual trace. The definition refines the Halpern-Pearl counterfactual-dependence clause by restricting the admissible preventing interventions to those reachable by feasible procedures. Concretely, the *Feasible prevention* clause in Definition 11 asks for a witness trace σ such that $s_0 \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$ and $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \neg\varphi_B$. Its minimality condition is resource-indexed: among feasible preventing traces, we select those minimal with respect to the preorder induced by $\otimes$. A secondary tie-breaker, such as trace length, may be added among incomparable traces. This matches responsibility attribution settings in which the salient notion of a smallest change is resource cost rather than variable-set cardinality.

4.4. Worked Example: k-of-n Pause Switch

We use a k-of-n approval (multisig) pause switch [34, 35] as a running example of a coordination-dependent alternative. A system can be paused only after approvals from at least k of n designated parties. No single signer can unilaterally prevent some outcome (called loss), but a sufficiently large coalition can do so exactly when the procedural constraints remain open.

We fix $n \geq 2$ and a threshold $k \in \{2, \dots, n\}$. Let E denote that an exploit attempt is underway, let W denote that the pause proposal window is still open, let $A_1, \dots, A_n$ denote approvals by signers, let Q denote that quorum is reached, let P denote that the system is paused in time, and let B denote the loss event.

Let $U_E, U_W, U_{A_1}, \dots, U_{A_n}$ be Boolean exogenous variables. Consider the deterministic acyclic SCM over endogenous variables $E, W, A_1, \dots, A_n, Q, P, B$ with equations

$$\begin{aligned}
&E := U_E, && W := U_W,\\
&A_i := U_{A_i} && \text{for each } i \in \{1, \dots, n\},\\
&Q := f_Q(A_1, \dots, A_n) && f_Q(a_1, \dots, a_n) = \mathbf{1}\left[\textstyle\sum_{i=1}^{n} a_i \geq k\right],\\
&P := f_P(E, Q) && f_P(e, q) = \mathbf{1}[\, e = 1 \wedge q = 1 \,],\\
&B := f_B(E, P) && f_B(e, p) = \mathbf{1}[\, e = 1 \wedge p = 0 \,].
\end{aligned}$$

Equivalently, since all variables are Boolean, one may write $P := E \wedge Q$ and $B := E \wedge \neg P$ as shorthand for the corresponding deterministic functions. Assume the actual context $\vec{u}$ satisfies $U_E = 1$, $U_W = 1$, and $U_{A_i} = 0$ for all i. Hence $(\mathcal{M}, \vec{u}) \models E{=}1$, $W{=}1$, and $A_i{=}0$ for all i, and therefore $Q{=}0$, $P{=}0$, and $B{=}1$.

For each signer $i \in \{1, \dots, n\}$, we introduce an agent i with $\mathcal{C}_i = \{A_i\}$: signer i can submit, at most once, an approval action $\alpha^i := (i, A_i{:=}1)$. Unconstrained counterfactual reasoning quantifies over *all* endogenous variables, so one may intervene on variables such as P or Q. In our running context $\vec{u}$ we have $(\mathcal{M}, \vec{u}) \models E{=}1$ and hence $(\mathcal{M}, \vec{u}) \models B{=}1$. Under an intervention that flips P, the loss is prevented $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(P{=}1)]\,(B{=}0)$.

Moreover, under the same standing assumption $(\mathcal{M}, \vec{u}) \models E{=}1$, flipping Q also prevents loss, since $P := f_P(E, Q)$ yields $P{=}1$ when $E{=}1$ and $Q{=}1$ $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(Q{=}1)]\,(B{=}0)$. These interventions are legitimate SCM interventions, but they do not represent executable options: neither P nor Q is controlled by any signer.

For each signer i, assume the observation map reveals whether the pause window is open $\mathsf{obs}_i(\vec{v}) = \vec{v}(W)$. Thus the observation space is $\mathcal{O}_i = \{0, 1\}$, and $o_i = 1$ means that signer i observes the pause proposal as still live. The approval action is enabled exactly when the signer observes the window to be open and has sufficient resources: $G_{\alpha^i}(\vec{u}, o_i, \rho, I) = 1$ iff $o_i = 1$ and $\mathsf{cost}(\alpha^i) \preceq \rho(i)$. Here $o_i = \mathsf{obs}_i(\vec{v}_I)$ and $\vec{v}_I = \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$. Let the initial state be $s = (\vec{u}, \rho, \emptyset)$. Since $U_W = 1$ in the running context, the initial valuation $\vec{v}_0 := \mathrm{Sol}_{\mathcal{M}}(\vec{u})$ satisfies $\vec{v}_0(W) = 1$.

We fix a signer i. Under the set-once discipline, an i-trace can write only A_i, and it can do so at most once. Thus in any intervened model reachable by an executable i-trace we have $\sum_{j=1}^{n} A_j \leq 1$, hence $Q = 0$, $P = 0$, and therefore $B = 1$ whenever $k > 1$. Consequently, for $k > 1$, $(\widehat{\mathcal{M}}, s) \nVdash \langle i \rangle \neg B$. This formalizes the intended no-unilateral-pause property: the loss outcome is not preventable by any single signer. Now let $C \subseteq \{1, \ldots, n\}$ be a coalition with $|C| \geq k$. Assume each signer in C has sufficient resources for a single approval.

Since no action writes W, and since $U_W = 1$ in the running context, every gate check along the relevant execution observes $o_i = 1$. Let us pick distinct $i_1, \ldots, i_k \in C$ and consider the set-once coalition trace $\sigma_C := (i_1, A_{i_1} := 1) ; \cdots ; (i_k, A_{i_k} := 1)$. If σ_C is executable from s, then in the resulting intervened model we have $\sum_{j=1}^{n} A_j \geq k$, hence $Q = 1$, $P = 1$ since $E = 1$, and therefore $B = 0$. Thus $(\widehat{\mathcal{M}}, s) \vDash \langle C \rangle \neg B$. If instead the context satisfies $U_W = 0$, then every induced valuation has $W = 0$, every signer observes $o_i = 0$, and no approval action is enabled. Hence $(\widehat{\mathcal{M}}, s) \nVdash \langle C \rangle \neg B$ for every coalition C.

5. Metatheory: Basic Properties

This section establishes the basic metatheory for the feasible-counterfactual semantics. Lemma 1 validates that trace evaluation is coherent (acyclic SCMs yield unique valuations under accumulated interventions) and, crucially, that the *set-once* discipline induces an intrinsic horizon bound: an agent or coalition cannot act more times than there are unwritten controlled variables available. Lemma 2 confirms that the coalitional modality behaves monotonically: enlarging a coalition cannot reduce its feasible options. Lemma 3 and Corollary 1 establish a robustness principle: enlarging an agent's resource endowment cannot eliminate an option that was already feasible. Theorem 1 is a conservativity result. Under the trivial-feasibility conditions for agent i, where every i-action is enabled whenever its target variable is still unwritten and all such actions have neutral cost, feasible achievement corresponds to existence of an ordinary SCM intervention over a subset of endogenous variables controlled by i. Finally, we conclude with an NP-complete result for bounded-horizon feasibility (Theorem 2), which reflects that witnesses are short and checkable. Throughout we fix a deterministic acyclic R-SCM $\widehat{\mathcal{M}}$ as in Section 4.

Lemma 1. *Let $s = (\vec{u}, \rho, I)$ be a state.*

1. $\mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ *exists and is unique.*
2. *If $s \xrightarrow{\sigma} s_m$ for some final state $s_m = (\vec{u}, \rho_m, I_m)$, then I_m is uniquely determined by s and σ.*
3. *If σ is an executable i-trace from s (i.e., every action in σ is of the form $(i, X := x)$), then $|\sigma| \leq |\mathcal{C}_i \setminus \mathrm{dom}(I)|$. More generally, if σ is an executable C-trace from s then $|\sigma| \leq |\mathcal{C}_C \setminus \mathrm{dom}(I)|$ where $\mathcal{C}_C := \bigcup_{j \in C} \mathcal{C}_j$.*

Proof sketch. By assumption, $\mathcal{M}$ is deterministic and acyclic, and so, $\mathcal{M}_I$ preserves acyclicity and determinism, and admits a unique solution in $\vec{u}$. Let us write $\sigma = \alpha_1; \cdots; \alpha_m$ with $\alpha_t = (j_t, X_t := x_t)$. By the set-once discipline condition in Definition 5, each step satisfies $X_t \notin \mathrm{dom}(I_{t-1})$. Hence the cumulative update $I_t := I_{t-1} \cup \{X_t \mapsto x_t\}$ is well-defined and depends only on the previous map and α_t. It can be shown, by induction on t, that I_t is uniquely determined by the initial I and the prefix $\alpha_1 ; \cdots ; \alpha_t$, and therefore $I_\sigma = I_m$ is uniquely determined by s and σ.

Let σ be an executable i-trace from $s = (\vec{u}, \rho, I)$. Each action in σ sets some $X \in \mathcal{C}_i$, and by set-once condition we never write the same controlled variable twice and never write a variable already in $\mathrm{dom}(I)$. Thus there exists an injection from the set of actions in σ to the set $\mathcal{C}_i \setminus \mathrm{dom}(I)$, and so, $|\sigma| \leq |\mathcal{C}_i \setminus \mathrm{dom}(I)|$. The coalitional case is identical, with $\mathcal{C}_C := \bigcup_{j \in C} \mathcal{C}_j$. □

Lemma 2 (Coalition monotonicity)**.** *Let $C \subseteq C' \subseteq \mathcal{A}$. For any state s and event formula φ, if $(\widehat{\mathcal{M}}, s) \vDash \langle C \rangle \varphi$ then $(\widehat{\mathcal{M}}, s) \vDash \langle C' \rangle \varphi$.*

Proof. Assume $(\widehat{\mathcal{M}}, s) \models \langle C \rangle \varphi$. Then there exists an executable trace $\sigma = \alpha_1 ; \cdots ; \alpha_m$ from s such that each $\alpha_t = (j_t, X_t := x_t)$ has $j_t \in C$ and, for some final state $s \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$, we have $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. Since $C \subseteq C'$, every acting agent j_t also lies in C', so the same σ is a C'-trace and witnesses $(\widehat{\mathcal{M}}, s) \models \langle C' \rangle \varphi$. □

To show that if an alternative was feasible for an agent, then granting them additional budget or time or privilege should not invalidate it, we compare states that agree on context and accumulated interventions, and differ only in the acting agent's resource component.

Definition 12 (Monotone gates)**.** *Fix an agent i. For stores $\rho, \rho' : \mathcal{A} \to \mathcal{T}$, write $\rho \preceq_i \rho'$ iff $\rho(j) = \rho'(j)$ for all $j \neq i$ and $\rho(i) \preceq \rho'(i)$, where $\preceq$ is the availability preorder induced by $\otimes$ (Definition 1). For states $s = (\vec{u}, \rho, I)$ and $s' = (\vec{u}, \rho', I)$ with the same $(\vec{u}, I)$, write $s \preceq_i s'$ iff $\rho \preceq_i \rho'$. We say the gates are* monotone in the acting agent's resources *if for every $\alpha = (i, X := x) \in \mathrm{Act}$ and all states $s \preceq_i s'$, letting $\vec{v} := \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ and $o := \mathsf{obs}_i(\vec{v})$, we have, if $G_\alpha(\vec{u}, o, \rho, I) = 1$ then $G_\alpha(\vec{u}, o, \rho', I) = 1$.* □

Lemma 3. *Assume gates are monotone in the acting agent's resources (Definition 12). Fix an agent i and states $s = (\vec{u}, \rho, I)$ and $s' = (\vec{u}, \rho', I)$ with $s \preceq_i s'$. Let $\sigma = \alpha_1; \cdots ; \alpha_m$ be an i-trace (i.e., each $\alpha_t = (i, X_t := x_t)$). If $s \xrightarrow{\sigma} (\vec{u}, \eta, I_\sigma)$, then there exists a store η' such that $s' \xrightarrow{\sigma} (\vec{u}, \eta', I_\sigma)$.*

Proof. We prove the claim by induction on the length $m := |\sigma|$. The base case is when $m = 0$. Then $\sigma = \epsilon$ and $s \xrightarrow{\epsilon} s$. So $I_\sigma = I$ and we may take $\eta' = \rho'$, giving $s' \xrightarrow{\epsilon} (\vec{u}, \rho', I) = (\vec{u}, \eta', I_\sigma)$.

The inductive step is as follows. Write $\sigma = \alpha; \tau$ where $\alpha = (i, X := x)$ and τ is an i-trace. Assume $s \xrightarrow{\alpha} s_1$ and $s_1 \xrightarrow{\tau} (\vec{u}, \eta, I_\sigma)$, where $s_1 = (\vec{u}, \rho_1, I_1)$. By Definition 5, we show that α is also executable from s'. Since σ is a trace, $\alpha \in \mathrm{Act}$, and since $s \preceq_i s'$, we have $\rho(j) = \rho'(j)$ for all $j \neq i$ and $\rho(i) \preceq \rho'(i)$. Hence there exists $d \in \mathcal{T}$ such that $\rho(i) \otimes d \downarrow$ and $\rho'(i) = \rho(i) \otimes d$. Let $c := \mathsf{cost}(\alpha)$. Because α is executable from s, there exists $r \in \mathcal{T}$ such that $c \otimes r \downarrow$ and $\rho(i) = c \otimes r$. Therefore $\rho'(i) = \rho(i) \otimes d = (c \otimes r) \otimes d$. Since $(c \otimes r) \otimes d$ is defined, associativity with definedness produces $r \otimes d \downarrow$, $c \otimes (r \otimes d) \downarrow$, $(c \otimes r) \otimes d = c \otimes (r \otimes d)$. Thus $\rho'(i) = c \otimes (r \otimes d)$. So $c \preceq \rho'(i)$, with residual $r' := r \otimes d$. Hence α's resource cost is payable from $\rho'(i)$.

Let $\vec{v} := \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ and $o := \mathsf{obs}_i(\vec{v})$. Since $s \xrightarrow{\alpha} s_1$, Definition 5 gives $X \notin \mathrm{dom}(I)$ and $G_\alpha(\vec{u}, o, \rho, I) = 1$. Because s and s' have the same intervention map I, the condition $X \notin \mathrm{dom}(I)$ also holds for s'. Moreover, by monotonicity of gates, $G_\alpha(\vec{u}, o, \rho, I) = 1$ implies $G_\alpha(\vec{u}, o, \rho', I) = 1$. Hence $G_\alpha(\vec{u}, o, \rho', I) = 1$. Therefore, by Definition 5, there exists a successor state $s'_1 = (\vec{u}, \rho'_1, I_1)$ with $I_1 = I \cup \{X \mapsto x\}$, $\rho'_1(i) = r'$, and $\rho'_1(j) = \rho'(j)$ for all $j \neq i$.

Moreover, the original execution of α from s leaves residual r for agent i, so $\rho_1(i) = r$. The execution of α from s' leaves residual $r' := r \otimes d$, so $\rho'_1(i) = r'$. Since $r \otimes d \downarrow$ and $r' = r \otimes d$, we have $r \preceq r'$. Hence $\rho_1(i) \preceq \rho'_1(i)$. For every $j \neq i$, the one-step execution of α leaves agent j's resource component unchanged. Hence $\rho_1(j) = \rho(j)$ and $\rho'_1(j) = \rho'(j)$. Since $s \preceq_i s'$, we have $\rho(j) = \rho'(j)$ for all $j \neq i$. Therefore $\rho_1(j) = \rho'_1(j)$ for all $j \neq i$. The successor states s_1 and s'_1 also have the same context $\vec{u}$ and the same intervention map I_1. Therefore $s_1 \preceq_i s'_1$.

Now we apply the induction hypothesis to the suffix trace τ from $s_1 \preceq_i s'_1$. Since $s_1 \xrightarrow{\tau} (\vec{u}, \eta, I_\sigma)$, there exists a store η' such that $s'_1 \xrightarrow{\tau} (\vec{u}, \eta', I_\sigma)$. Finally, composing the step for α with the execution of τ gives $s' \xrightarrow{\alpha} s'_1 \xrightarrow{\tau} (\vec{u}, \eta', I_\sigma)$, i.e., $s' \xrightarrow{\sigma} (\vec{u}, \eta', I_\sigma)$ as required. □

Corollary 1. *Assume gates are monotone in the acting agent's resources (Definition 12). Let $s \preceq_i s'$. For every event formula φ, if $(\widehat{\mathcal{M}}, s) \models \langle i \rangle \varphi$, then $(\widehat{\mathcal{M}}, s') \models \langle i \rangle \varphi$.*

Proof. Assume $(\widehat{\mathcal{M}}, s) \models \langle i \rangle \varphi$. Then there exists an i-trace σ such that $s \xrightarrow{\sigma} (\vec{u}, \eta, I_\sigma)$ and $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. By Lemma 3, since $s \preceq_i s'$ there exists η' with $s' \xrightarrow{\sigma} (\vec{u}, \eta', I_\sigma)$. It follows that $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. Hence $(\widehat{\mathcal{M}}, s') \models \langle i \rangle \varphi$. □

The following theorem shows that when feasibility is trivial for agent i, the feasible-counterfactual modality collapses to ordinary SCM intervention over endogenous variables that are controlled by i. Under trivial feasibility for i, the formula $\langle i \rangle \varphi$ holds exactly when φ is true under some intervention map J with $\mathrm{dom}(J) \subseteq \mathcal{C}_i$.

Theorem 1. *Assume $\widehat{\mathcal{M}}$ satisfies trivial feasibility (Definition 9) for agent i. Then for every state $s = (\vec{u}, \rho, \emptyset)$ and every event formula φ, $(\widehat{\mathcal{M}}, s) \models \langle i \rangle \varphi$ iff there exists J with $\mathrm{dom}(J) \subseteq C_i$ such that $(\mathcal{M}_J, \vec{u}) \models \varphi$. Moreover, any such J can be witnessed by an executable trace that assigns each $X \in \mathrm{dom}(J)$ exactly once.*

Proof. Let us fix $s = (\vec{u}, \rho, \emptyset)$ and an event formula φ. Assume $(\widehat{\mathcal{M}}, s) \models \langle i \rangle \varphi$. By Definition 8, there exists an executable trace $\sigma = \alpha_1 ; \cdots ; \alpha_m$ from s such that every α_t has the form $(i, X_t := x_t)$ with $X_t \in C_i$, and, for some final state $s \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$, we have $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. Let $J := I_\sigma$. Therefore, every assignment recorded in J targets a variable in C_i, hence $\mathrm{dom}(J) \subseteq C_i$, and so, J is an i-control intervention. Moreover, we have $(\mathcal{M}_J, \vec{u}) \models \varphi$. This proves the forward implication.

In the other direction, assume there exists an intervention map J with $\mathrm{dom}(J) \subseteq C_i$ such that $(\mathcal{M}_J, \vec{u}) \models \varphi$. Let us enumerate $\mathrm{dom}(J) = \{X_1, \dots, X_m\}$ (with no repetition) and define the trace $\sigma := (i, X_1 := J(X_1)) ; \cdots ; (i, X_m := J(X_m))$. We show that σ is executable from s and that its cumulative intervention map is $I_\sigma = J$.

Define $I_0 := \emptyset$ and $I_t := I_{t-1} \cup \{X_t \mapsto J(X_t)\}$ for $t = 1, \dots, m$ as in Definition 2. Since all X_t are distinct and $I_0 = \emptyset$, we have $X_t \notin \mathrm{dom}(I_{t-1})$ at each step, so set-once discipline is respected. We now prove by induction on $t \in \{1, \dots, m\}$ that the t-th action is executable from the state reached after the prefix of length $t-1$. Let $s_{t-1} = (\vec{u}, \rho_{t-1}, I_{t-1})$ be the state reached after executing the first $t-1$ actions (with $s_0 = s$). Consider $\alpha_t = (i, X_t := J(X_t))$. By trivial feasibility, since $X_t \in C_i$ and $J(X_t) \in \mathcal{R}(X_t)$, we have $\alpha_t \in \mathrm{Act}$ and $\mathsf{cost}(\alpha_t) = e$. To apply the one-step rule (Definition 5) it suffices to check that $X_t \notin \mathrm{dom}(I_{t-1})$ and that $G_{\alpha_t}(\vec{u}, o_{t-1}, \rho_{t-1}, I_{t-1}) = 1$. The first condition holds by construction.

Let $\vec{v}_{t-1} := \mathrm{Sol}_{\mathcal{M}_{I_{t-1}}}(\vec{u})$ and $o_{t-1} := \mathsf{obs}_i(\vec{v}_{t-1})$ as in Definition 5. By Definition 9 applied to i in $\widehat{\mathcal{M}}$, and since $X_t \notin \mathrm{dom}(I_{t-1})$, we have $G_{\alpha_t}(\vec{u}, o_{t-1}, \rho_{t-1}, I_{t-1}) = 1$. By Definition 1, $e \otimes \rho_{t-1}(i)$ is defined and equals $\rho_{t-1}(i)$. Hence taking $r_t := \rho_{t-1}(i)$ suffices, and the updated store satisfies $\rho_t(i) = r_t = \rho_{t-1}(i)$, while $\rho_t(j) = \rho_{t-1}(j)$ for $j \neq i$. Thus $s_{t-1} \xrightarrow{\alpha_t} s_t$ for some $s_t = (\vec{u}, \rho_t, I_t)$ with $I_t = I_{t-1} \cup \{X_t \mapsto J(X_t)\}$. Therefore σ is executable from s, i.e., $s \xrightarrow{\sigma} (\vec{u}, \rho_m, I_m)$.

It remains to identify the final intervention map. By construction, I_m assigns exactly the variables in $\{X_1, \dots, X_m\} = \mathrm{dom}(J)$, and for each such variable X_t we have $I_m(X_t) = J(X_t)$. Hence we can take $I_\sigma = I_m = J$. Finally, since $(\mathcal{M}_J, \vec{u}) \models \varphi$, we have $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. By the definition of the Feasible-counterfactual modality (Definition 8), this witnesses $(\widehat{\mathcal{M}}, s) \models \langle i \rangle \varphi$. □

Definition 13 (Bounded-horizon feasibility)**.** *Let φ be an event formula, let $C \subseteq \mathcal{A}$ be a coalition, and let $k \in \mathbb{N}$ be given in unary. The problem* BoundedFeasibility *asks whether, from the initial state $s_0 = (\vec{u}, \rho, \emptyset)$, there exists a C-trace $\sigma = \alpha_1 ; \cdots ; \alpha_m$ with $m \leq k$ such that σ is executable from s_0 and, for some final state $s_0 \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$, we have $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$.* □

We specify k in unary in the bounded-horizon reachability problem: a YES certificate is a trace of length at most k, so unary k makes the certificate size polynomial in the input. The interested reader may consult Chapter 5 of [6] and [36, 37] for a discussion of computational complexity results for the Halpern-Pearl framework. We follow [36] for proving Theorem 2.

Theorem 2 (Bounded-horizon feasibility is NP-complete)**.** *Assume that all endogenous variables have finite domains and that $\widehat{\mathcal{M}}$ is acyclic. Assume that R-SCM instances are given by finite, polynomial-size descriptions. In particular, variable domains, actions, resource objects, stores, contexts, observation values, and event formulae have polynomial-size encodings. Membership in* Act *and in the control sets C_i is decidable in polynomial time, and the structural functions, observation maps, cost function, gate predicates, resource-composition checks, and event-formula evaluation are all polynomial-time computable under these encodings. Given $c, r, t \in \mathcal{T}$, one can decide whether $c \otimes r \downarrow$ and whether $c \otimes r = t$. With k given in unary,* BoundedFeasibility *is NP-complete.* □

Proof. We first show that BoundedFeasibility $\in$ NP. A nondeterministic certificate consists of $\sigma = \alpha_1 ; \cdots ; \alpha_m$, and $\rho_0, \rho_1, \dots, \rho_m$ where $m \leq k$, $\rho_0 = \rho$, and each ρ_t is a resource store. Since k is given in unary and all action names and resource objects have polynomial-size encodings, this certificate has

polynomial size. The verifier checks executability step by step. It first checks that every ρ_t is a well-formed encoded store assigning a resource object to each agent. Set $I_0 := \emptyset$. For each $t = 0, \dots, m-1$, write $\alpha_{t+1} = (i_{t+1}, X_{t+1} := x_{t+1})$. The verifier first checks that $\alpha_{t+1} \in \mathrm{Act}$, that $i_{t+1} \in C$, that $X_{t+1} \in C_{i_{t+1}}$, that $x_{t+1} \in \mathcal{R}(X_{t+1})$, and that $X_{t+1} \notin \mathrm{dom}(I_t)$. It then computes the unique valuation $\vec{v}_t = \mathrm{Sol}_{\mathcal{M}_{I_t}}(\vec{u})$ by evaluating the acyclic structural equations in topological order. This is polynomial time because the SCM is acyclic and each f_V is polynomial-time evaluable. It computes $o_t := \mathsf{obs}_{i_{t+1}}(\vec{v}_t)$ and checks the gate condition $G_{\alpha_{t+1}}(\vec{u}, o_t, \rho_t, I_t) = 1$. It remains to verify that the resource update is legitimate. Let $c_{t+1} := \mathsf{cost}(\alpha_{t+1})$. The verifier checks that all non-acting agents' stores are unchanged $\rho_{t+1}(j) = \rho_t(j)$ for all $j \neq i_{t+1}$. It then checks that the acting agent's next store is a valid residual $c_{t+1} \otimes \rho_{t+1}(i_{t+1}) \downarrow$ and $c_{t+1} \otimes \rho_{t+1}(i_{t+1}) = \rho_t(i_{t+1})$. This is polynomial time by assumption. Also, non-uniqueness of residuals is harmless because the residual store is part of the certificate. Finally, the verifier updates $I_{t+1} := I_t \cup \{X_{t+1} \mapsto x_{t+1}\}$. After all m steps, it computes $\vec{v}_m = \mathrm{Sol}_{\mathcal{M}_{I_m}}(\vec{u})$ and checks whether $\vec{v}_m \models \varphi$, equivalently whether $(\mathcal{M}_{I_m}, \vec{u}) \models \varphi$. All checks are polynomial time, hence BoundedFeasibility $\in$ NP.

For NP-hardness, we reduce SAT to BoundedFeasibility. Let ψ be a CNF formula over propositional variables $x_1, \dots, x_n$. We construct, in polynomial time, an instance of BoundedFeasibility that is a YES-instance iff ψ is satisfiable. Define a Boolean acyclic SCM with endogenous variables $X_1, \dots, X_n, Y$ and ranges $\mathcal{R}(X_j) = \mathcal{R}(Y) = \{0, 1\}$. We use the unique empty context $\vec{u}_\emptyset$. For each j, set $X_j := 0$, and $Y := \psi(X_1, \dots, X_n)$, where ψ is evaluated as the Boolean function obtained by interpreting each propositional variable x_j as the endogenous variable X_j. The graph is acyclic, since each X_j has no endogenous parents and Y depends only on $X_1, \dots, X_n$. Now add one agent i with $C_i = \{X_1, \dots, X_n\}$. For every $j \in \{1, \dots, n\}$ and $b \in \{0, 1\}$, include the primitive action $(i, X_j := b)$ in Act. Let the coalition be $C = \{i\}$.

We use the trivial one-object resource frame $\mathcal{T} = \{e\}$, and $e \otimes e = e$. Set $\mathsf{cost}(\alpha) = e$ for every action α, and take the initial store $\rho(i) = e$. Let $\mathcal{O}_i = \{o_0\}$ and define the observation map by $\mathsf{obs}_i(\vec{v}) = o_0$ for all valuations $\vec{v}$. Define every gate to be always open $G_\alpha(\vec{u}, o_0, \rho, I) = 1$ for all contexts $\vec{u}$, stores ρ, and cumulative intervention maps I. The set-once condition is still enforced by the one-step execution rule.

The initial state is therefore the full state triple $s_0 = (\vec{u}_\emptyset, \rho, \emptyset)$. Set the goal formula to $\varphi \equiv (Y = 1)$, and set the horizon to $k := n$. Suppose first that ψ is satisfiable. Let $a_1, \dots, a_n \in \{0, 1\}$ be a satisfying assignment. Consider the trace $\sigma := (i, X_1 := a_1); \cdots; (i, X_n := a_n)$. Every step is executable as each target variable is controlled by i, each variable is written at most once, every gate is open, and every cost is neutral. The final cumulative intervention map satisfies $I_\sigma(X_j) = a_j$ for every j. Hence, in the intervened model $\mathcal{M}_{I_\sigma}$, $Y = \psi(a_1, \dots, a_n) = 1$. Thus $(\mathcal{M}_{I_\sigma}, \vec{u}_\emptyset) \models \varphi$, so the constructed feasibility instance is a YES-instance.

Conversely, suppose the constructed feasibility instance is a YES-instance. Then there is an executable C-trace σ of length at most n such that, for its final cumulative intervention map I_σ, $(\mathcal{M}_{I_\sigma}, \vec{u}_\emptyset) \models (Y = 1)$. Define a Boolean assignment $a_1, \dots, a_n$ by $a_j := I_\sigma(X_j)$, if $X_j \in \mathrm{dom}(I_\sigma)$, and $a_j := 0$. This is well-defined because the set-once condition prevents any X_j from being assigned twice along an executable trace. For every j, the value of X_j in $\mathrm{Sol}_{\mathcal{M}_{I_\sigma}}(\vec{u}_\emptyset)$ is exactly a_j. Here, intervened variables take their intervened values, and unintervened variables retain their structural default value 0. Since $Y := \psi(X_1, \dots, X_n)$ and the final valuation satisfies $Y = 1$, we obtain $\psi(a_1, \dots, a_n) = 1$. Thus ψ is satisfiable.

The construction is polynomial in $|\psi|$, and $k = n$ is unary-encoded. Therefore SAT $\leq_p$ BoundedFeasibility, so BoundedFeasibility is NP-hard. Together with membership in NP, this proves that BoundedFeasibility is NP-complete. □

6. Conclusion

We have characterized an *intervention–feasibility gap* in the Halpern-Pearl (HP) account of actual causation. The proposed R-SCMs (Definition 7) address this by preserving standard SCM intervention semantics while restricting agent-relative alternatives. The resulting semantics separates causal dynamics from procedural feasibility. The dynamics layer remains ordinary Halpern-Pearl evaluation in the intervened model $\mathcal{M}_I$, while the feasibility layer is a labelled transition system over states $(\vec{u}, \rho, I)$. We

focus on deterministic acyclic SCMs and bounded traces to keep counterfactual evaluation unambiguous and obtain crisp complexity bounds, in particular, NP-completeness of bounded-horizon feasibility.

The resource component is substructural in a minimal semantic sense: feasibility is tracked in a partial commutative monoid $(\mathcal{T}, \otimes, e)$, so resources compose and exclusivity is enforced by the partiality of $\otimes$ which is closely aligned with the partial-commutative-monoid semantics used to model disjointness in BI (the logic of bunched implications) and separation-logic semantics [38, 39, 40, 41].

In consequence, $(\widehat{\mathcal{M}}, s) \models \langle C \rangle \varphi$ is witnessed by an explicit executable trace σ such that $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. We hypothesize that a BI-inspired system [38, 40, 41] can internalize this as a two-context judgment with a structural fact context and a substructural resource store. It also suggests a lightweight tooling path in which feasibility and preventability claims are accompanied by machine-checkable certificates that can be verified independently as auditable artifacts or synthesized via SAT/SMT back-ends. This aligns with interface-level systems modelling where procedures and controls are viewed as explicit design elements [42, 43]. Other possible extensions include richer resource algebras for shared or global exclusives, non-commutative composition for ordered workflows, and epistemic or probabilistic feasibility constraints.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-5.5 via Codex with xhigh mode for Grammar and spelling check, Formatting assistance, and Improve writing style. After using this tool/service, the author(s) reviewed and further edited the content as needed and take full responsibility for the publication's content.

References

[1] H. A. Simon, A Behavioral Model of Rational Choice, The Quarterly Journal of Economics 69 (1955) 99–118. doi:10.2307/1884852.

[2] H. A. Simon, The Sciences of the Artificial, Reissue of the 3rd ed. ed., The MIT Press, 2019. doi:10.7551/mitpress/12107.001.0001.

[3] M. S. Feldman, B. T. Pentland, Reconceptualizing Organizational Routines as a Source of Flexibility and Change, Administrative Science Quarterly 48 (2003) 94–118. doi:10.2307/3556620.

[4] F. S. de Sio, J. van den Hoven, Meaningful Human Control over Autonomous Systems: A Philosophical Account, Frontiers in Robotics and AI 5 (2018). URL: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2018.00015. doi:10.3389/frobt.2018.00015.

[5] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed., Cambridge University Press, 2009.

[6] J. Y. Halpern, Actual Causality, The MIT Press, Cambridge, MA, 2016. doi:10.7551/mitpress/10809.001.0001.

[7] H. Chockler, J. Y. Halpern, Responsibility and Blame: A Structural-Model Approach, Journal of Artificial Intelligence Research 22 (2004) 93–115. doi:10.1613/jair.1391.

[8] B. Ustun, A. Spangher, Y. Liu, Actionable Recourse in Linear Classification, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19, Association for Computing Machinery, 2019, pp. 10–19. doi:10.1145/3287560.3287566.

[9] R. Poyiadzi, K. Sokol, R. Santos-Rodriguez, T. D. Bie, P. Flach, FACE: Feasible and Actionable Counterfactual Explanations, in: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, AIES '20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 344–350. doi:10.1145/3375627.3375850.

[10] A.-H. Karimi, B. Schölkopf, I. Valera, Algorithmic Recourse: from Counterfactual Explanations to Interventions, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 353–362. doi:10.1145/3442188.3445899.

[11] N. Tran, C. Baral, Encoding Probabilistic Causal Model in Probabilistic Action Language, in: Proceedings of the 19th National Conference on Artificial Intelligence, AAAI'04, AAAI Press, 2004, pp. 305–310.
[12] M. Hopkins, J. Pearl, Causality and Counterfactuals in the Situation Calculus, Journal of Logic and Computation 17 (2007) 939–953. doi:10.1093/logcom/exm048.
[13] V. Batusov, M. Soutchanski, Situation Calculus Semantics for Actual Causality, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18, AAAI Press, 2018, pp. 1744–1752. doi:10.1609/aaai.v32i1.11561.
[14] A. Finzi, T. Lukasiewicz, Structure-Based Causes and Explanations in the Independent Choice Logic, in: C. Meek, U. Kjærulff (Eds.), Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence, UAI'03, Morgan Kaufmann, 2003, pp. 225–232.
[15] D. Liu, V. Belle, What Is a Counterfactual Cause in Action Theories?, in: Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '25, International Foundation for Autonomous Agents and Multiagent Systems, 2025, pp. 2627–2629.
[16] E. LeBlanc, M. Balduccini, J. Vennekens, Explaining Actual Causation via Reasoning About Actions and Change, in: Logics in Artificial Intelligence, volume 11468 of *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 231–246. doi:10.1007/978-3-030-19570-0_15.
[17] D. Galles, J. Pearl, Axioms of Causal Relevance, Artificial Intelligence 97 (1997) 9–43. doi:10.1016/S0004-3702(97)00047-7.
[18] J. Y. Halpern, M. Kleiman-Weiner, Towards formal definitions of blameworthiness, intention, and moral responsibility, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18, AAAI Press, 2018, pp. 1853–1860. doi:10.1609/aaai.v32i1.11557.
[19] S. Beckers, J. Y. Halpern, C. Hitchcock, Causal Models with Constraints, in: M. van der Schaar, C. Zhang, D. Janzing (Eds.), Proceedings of the Second Conference on Causal Learning and Reasoning, volume 213 of *Proceedings of Machine Learning Research*, PMLR, 2023, pp. 866–879. URL: https://proceedings.mlr.press/v213/beckers23a.html.
[20] J. Y. Halpern, C. Hitchcock, Graded Causation and Defaults, The British Journal for the Philosophy of Science 66 (2015) 413–457. doi:10.1093/bjps/axt050.
[21] C. Baral, M. Gelfond, Reasoning Agents in Dynamic Domains, in: Logic-Based Artificial Intelligence, Springer US, Boston, MA, 2000, pp. 257–279. doi:10.1007/978-1-4615-1567-8_12.
[22] R. Reiter, Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems, The MIT Press, 2001. doi:10.7551/mitpress/4074.001.0001.
[23] L. Bertossi, Declarative Approaches to Counterfactual Explanations for Classification, Theory and Practice of Logic Programming 23 (2023) 559–593. doi:10.1017/S1471068421000582.
[24] Q. Shi, Responsibility in Extensive Form Games, Proceedings of the AAAI Conference on Artificial Intelligence 38 (2024) 19920–19928. doi:10.1609/aaai.v38i18.29968.
[25] T. Parker, U. Grandi, E. Lorini, Responsibility in a Multi-value Strategic Setting, in: Multi-Agent Systems: 21st European Conference, EUMAS 2024, Dublin, Ireland, August 26–28, 2024, Proceedings, volume 15685 of *Lecture Notes in Computer Science*, Springer Nature Switzerland, Cham, 2025, pp. 138–158. doi:10.1007/978-3-031-93930-3_9.
[26] E. Lorini, D. Longin, E. Mayor, A Logical Analysis of Responsibility Attribution: Emotions, Individuals and Collectives, Journal of Logic and Computation 24 (2014) 1313–1339. doi:10.1093/logcom/ext072.
[27] N. Alechina, B. Logan, N. H. Nga, A. Rakib, Resource-bounded alternating-time temporal logic, in: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1, AAMAS '10, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2010, pp. 481–488.
[28] H. N. Nguyen, A. Rakib, Formal Modelling and Verification of Probabilistic Resource Bounded

Agents, Journal of Logic, Language and Information 32 (2023) 829–859. doi:10.1007/s10849-023-09405-1.

[29] V. Yazdanpanah, M. Dastani, W. Jamroga, N. Alechina, B. Logan, Strategic Responsibility Under Imperfect Information, in: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 592–600.

[30] H. Wang, H. Zou, X. Zhou, S. Wang, W. Yang, P. Cui, Learning Feasible Causal Algorithmic Recourse: A Prior Structural Knowledge Free Approach, in: Proceedings of the ACM on Web Conference 2025, WWW '25, Association for Computing Machinery, New York, NY, USA, 2025, pp. 4507–4518. doi:10.1145/3696410.3714859.

[31] J. von Kügelgen, A.-H. Karimi, U. Bhatt, I. Valera, A. Weller, B. Schölkopf, On the Fairness of Causal Algorithmic Recourse, Proceedings of the AAAI Conference on Artificial Intelligence 36 (2022) 9584–9594. URL: https://ojs.aaai.org/index.php/AAAI/article/view/21192. doi:10.1609/aaai.v36i9.21192.

[32] J. Winn, Causality with Gates, in: N. D. Lawrence, M. Girolami (Eds.), Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, volume 22 of *Proceedings of Machine Learning Research*, PMLR, La Palma, Canary Islands, 2012, pp. 1314–1322. URL: https://proceedings.mlr.press/v22/winn12.html.

[33] F. Eberhardt, Introduction to the Epistemology of Causation, Philosophy Compass 4 (2009) 913–925. doi:10.1111/j.1747-9991.2009.00243.x.

[34] K. Itakura, K. Nakamura, A Public-key Cryptosystem Suitable for Digital Multisignatures, NEC Research & Development (1983) 1–8.

[35] S. Micali, K. Ohta, L. Reyzin, Accountable-subgroup Multisignatures: Extended Abstract, in: Proceedings of the 8th ACM Conference on Computer and Communications Security, Association for Computing Machinery, New York, NY, USA, 2001, pp. 245–254. doi:10.1145/501983.502017.

[36] T. Eiter, T. Lukasiewicz, Complexity results for structure-based causality, Artificial Intelligence 142 (2002) 53–89. doi:10.1016/S0004-3702(02)00271-0.

[37] G. Aleksandrowicz, H. Chockler, J. Y. Halpern, A. Ivrii, The Computational Complexity of Structure-Based Causality, Journal of Artificial Intelligence Research 58 (2017) 431–451. doi:10.1613/jair.5229.

[38] S. S. Ishtiaq, P. W. O'Hearn, BI as an Assertion Language for Mutable Data Structures, in: Proceedings of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, Association for Computing Machinery, 2001, pp. 14–26. doi:10.1145/360204.375719.

[39] J. C. Reynolds, Separation Logic: A Logic for Shared Mutable Data Structures, in: Proceedings of the 17th IEEE Symposium on Logic in Computer Science (LICS), IEEE, 2002, pp. 55–74. doi:10.1109/LICS.2002.1029817.

[40] D. Pym, Resource semantics: logic as a modelling technology, ACM SIGLOG News 6 (2019) 5–41. doi:10.1145/3326938.3326940.

[41] A. V. Gheorghiu, D. J. Pym, Semantical Analysis of the Logic of Bunched Implications, Studia Logica 111 (2023) 525–571. doi:10.1007/s11225-022-10028-z.

[42] P. Chakraborty, T. Caulfield, D. Pym, Local Causal Reasoning in Multiagent Systems (Extended Abstract), in: Advances in Explainable Agentic AI and Large Language Models, volume 2923 of *Communications in Computer and Information Science*, Springer, Cham, 2026, pp. 73–95. doi:10.1007/978-3-032-20548-3_6.

[43] P. Chakraborty, T. Caulfield, D. Pym, A Logic for Resource-Sensitive Coalition Games, in: Game Theory and AI for Security, volume 16223 of *Lecture Notes in Computer Science*, Springer, Cham, 2026, pp. 61–80. doi:10.1007/978-3-032-08064-6_4.