Feasible Counterfactual Reasoning for Responsibility and Recourse

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 modeled by a partial commutative monoid. We interpret agent-relative counterfactual feasibility via executable, resource-consuming intervention traces (a sequence of actions) 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.

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 are often subject to ex post audit (e.g., compliance checks, incident postmortems, or regulatory review). In such settings, whether an agent 'could have done otherwise' is rarely a procedural question about what was *enabled* for that agent. This viewpoint is classical in the bounded rationality tradition (Simon 1955; Simon 2019) and is reflected empirically in how organizations encode action through routines and interfaces (Feldman and Pentland 2003). It is also explicit in operational and regulatory practice. Human oversight in safety-critical automation is meaningful only insofar as operators have genuine, feasible intervention options (de Sio and van den Hoven 2018).

Counterfactual reasoning is the standard bridge from descriptive models of a system to normative questions about explanation, responsibility, and recourse. The dominant formal substrate for this is the Pearl-style structural causal models (SCMs) and the Halpern-Pearl (HP) account of actual causation and its graded variants (Pearl 2009; Halpern 2016; Chockler and Halpern 2003). SCMs employ structural equations to specify how endogenous variables are determined by exogenous context, while the intervention operator $\mathrm{do}(\cdot)$ supports counterfactual queries by replacing selected variables by constants. In this semantics, explaining (or assigning responsibility for) an outcome amounts to evaluating what would have happened under a hypothetical manipulation of some variables, holding the structural dynamics fixed (Pearl 2009; Halpern 2016). This treats counterfactual antecedents as freely specifiable variable-settings. We discuss in the next section why that idealization becomes problematic when 'could have done otherwise' is assessed against procedural constraints.

A recurring issue in many socio-technical scenarios is that the most salient alternatives are often not executable options the relevant agent actually had in the situation at hand. As mentioned, standard SCM semantics, treats interventions as a purely semantic operation on equations (Pearl 2009). For any endogenous variable X and any value $x \in \mathcal{R}(X)$, the intervention $\mathrm{do}(X{=}x)$ is admissible. This generality is built directly into the Halpern-Pearl (HP) account of actual causation in which counterfactual dependence is witnessed *under a contingency*. One is allowed to intervene not only on the putative set of cause variables X, but also on an auxiliary set W of endogenous variables, fixing them to a chosen setting $\vec{w}$, so as to make the effect φ counterfactually sensitive to changing X (the modified AC2 contingency clause in (Halpern 2016)). Because SCM semantics imposes no restriction on which endogenous variables may appear in such interventions, the resulting counterfactual language ranges over two qualitatively different kinds of antecedents: assignments that correspond to genuine control moves available to a particular agent in the actual context, and assignments that are physically or normatively impossible for that agent to realize.

The Halpern-Pearl formalism of actual causation ties explanation, blame, and graded responsibility to counterfactual dependence (Halpern 2016; Chockler and Halpern 2003;

Beckers, Halpern, and Hitchcock 2023), while algorithmic causal recourse methods search for outcome-flipping counterfactual changes to input features (Ustun, Spangher, and Liu 2019). As a result, responsibility claims and recourse recommendations that are mathematically well-defined yet procedurally infeasible are supported. We call this mismatch the *intervention–feasibility gap*. Closely related concerns arise in the recourse literature, which distinguishes valid counterfactual reasoning from actionable and feasible changes (Ustun, Spangher, and Liu 2019; Poyiadzi et al. 2020) and argues that recourse should be cast as (minimal) interventions in an underlying causal model rather than arbitrary feature-setting moves (Karimi, Scholkopf, and Valera 2021; Ustun, Spangher, and Liu 2019).

A natural response is to move to an action-theoretic representation in which feasibility is explicit: one specifies an action signature with preconditions (including timing, permissions, and coordination requirements), and counterfactual claims are evaluated against an explicit narrative of what actions were executable. The Knowledge Representation (KR) community has long studied causal reasoning in such formalisms. The interested reader may consult (Tran and Baral 2004; Hopkins and Pearl 2007; Batusov and Soutchanski 2018; Finzi and Lukasiewicz 2002; Liu and Belle 2025; LeBlanc, Balduccini, and Vennekens 2019) among many other works which we omit for the sake of brevity. In these settings, 'intervention' is largely aligned with the notion of action by construction, and so, 'could have done otherwise' is constrained syntactically. Although we derive inspiration for our theoretical apparatus from these approaches, our approach is deliberately *SCM-native*.

While an SCM can be compiled into an action-theoretic formalism, and as mentioned, the KR literature offers several bridges between structural models and action-oriented formalisms (Section 2), doing so typically introduces additional operational commitments that are largely orthogonal to our objective. This invites a natural comparison: are we simply doing deterministic planning (or an ability logic with costs) with SCM evaluation used as a simulator? The key difference is methodological, in that, we do not assume an action model ex ante.

Instead, we propose a conservative extension that preserves SCM substrate in HP-formalism intact, while making *admissibility* of counterfactual alternatives explicit. Our technical move is to layer a 'lightweight' procedural semantics on top of an SCM: an agent 'could have done otherwise' only if there exists an *executable trace* of enabled interventions available to that agent (or coalition) in the current situation. These agent-relative interventions operate only on a designated set of corresponding agent-relative endogenous variables. Whether an action is enabled is decided by a *gate* predicate that may depend on the current context, the agent's observation of the current valuation, and a consumable resource store (e.g., time, permissions, budget, quorum capacity). This results in a feasible-counterfactual modality $\langle i \rangle \varphi$ that ranges over action sequences the agent i can perform relative to resources available to it.

A further technical point is *resource sensitivity*. In our setting, feasibility is a state-dependent constraint driven by resources such as time, privilege, budget, or coordination capacity. Resource-bounded strategic logics and responsibility formalisms in multi-agent systems already internalize similar agency constraints at the semantic level (e.g., coalition ability, responsibility under imperfect information, and STIT-based accounts) (Alechina et al. 2010; Nguyen et al. 2019; Yazdanpanah et al. 2019; Shi 2024; Parker, Grandi, and Lorini 2025; Lorini, Longin, and Mayor 2013). Starting from an SCM, we use explicit resource constraints to prune the space of counterfactual antecedents relevant to responsibility and recourse.

The underlying SCM evaluation is left unchanged but the admissibility layer imposes a restriction: an intervention trace is admissible only if each step passes its enablement predicates (gates) and its cost can be paid from the current resource store. Consequently, resources cannot be duplicated or used 'for free', and exclusivity clashes (e.g., mutexes or quorum slots) are captured by the partiality of the resource composition. On this reading, feasibility is certified by an executable trace of primitive actions available to the agent in the current state whose accumulated interventions make the target event true. Minimality is then a claim that among all feasible traces which achieve the same outcome, the chosen witness is minimal with respect to the preorder induced by the resource algebra.

Our contributions are as follows. First, we characterize the *intervention–feasibility gap* in HP-style responsibility analyses in Section 1. Section 2 situates the approach relative to action-theoretic accounts of causation, logics of agency and responsibility, resource-bounded ability logics, and the causal recourse literature, clarifying how our contribution is a restriction principle for admissibility. In Section 3, we present minimal examples exposing the intervention–feasibility gap in responsibility and recourse. Section 4 introduces the R-SCM semantics. Section 5 formalizes the main reasoning task (bounded-horizon feasible achievement and prevention) and proves some basic properties of R-SCMs. Section 6 concludes with limitations and extensions, including a proof-certificate direction for auditable feasibility witnesses.

2 Related Work

Our starting point is Pearl's structural-equations account of causality (Pearl 2009; Galles and Pearl 1997), and its development through the Halpern-Pearl (HP) framework for actual causality and its variants (Halpern 2016). The agent-relative stakes considered here are closest in spirit to graded refinements such as responsibility and blame (Chockler and Halpern 2003; Halpern and Kleiman-Weiner 2018). Our contribution is orthogonal to refinements of the HP definition itself: we address a semantic mismatch that interventions range over all endogenous variables rather than over an agent's executable options. The first position is an extrinsic view of the underlying system whereas the later is an intrinsic agent-relative view.

In this direction, several works connect Pearl-style causal models to explicit action languages. Tran and Baral encode probabilistic causal models into the probabilistic action language PAL, motivated by the mismatch between

Pearl's intervention vocabulary and action-theoretic descriptions (Tran and Baral 2004). Finzi and Lukasiewicz integrate structural-model causality with Poole's independent choice logic (Poole 1997) to obtain a first-order setting with explicit actions while importing HP-style causality and explanation (Finzi and Lukasiewicz 2002). These results establish representational compatibility between SCMs and action formalisms, but they neither address the intervention–feasibility gap nor do they give an agent-dependent counterfactual admissibility criterion.

A substantial line of work studies counterfactual and actual causality directly in action-theoretic settings such as the situation calculus. Hopkins and Pearl investigate causality and counterfactual reasoning in the situation calculus, partly as a response to expressiveness limitations of structural models for dynamic domains (Hopkins and Pearl 2007). Batusov and Soutchanski rebuild actual causation inside atemporal situation calculus (Batusov and Soutchanski 2018). Their stated motive is to mitigate the lack of objective criteria to select admissible counterfactual possible worlds for analysis. They introduce a notion of achievement, and also give a translation from acyclic causal models to situation calculus and relate their notion to HP-style causality under their embedding. LeBlanc *et al.* propose an action-theoretic framework for explaining actual causation, formalized in the action language AL (Baral and Gelfond 2000) and implement the framework using Answer Set Programming (LeBlanc, Balduccini, and Vennekens 2019). Liu and Belle propose a counterfactual notion of achievement cause in action theory (Reiter 2001) and analyze its relationship to achievement causes in the situation calculus tradition (Liu and Belle 2025). These works are close in spirit to ours, and we acknowledge the methodological inspiration, in that they bind causation to action structure. However, they largely start from an *action description point of view* where feasibility is encoded by the given action signatures and preconditions. The novelty of our approach, instead, is a restriction principle for counterfactual reasoning relative to an SCM itself within the HP-framework.

Causal reasoning has also been studied through logical formalisms that connect to non-monotonic knowledge representation and default reasoning. For example, Bochman shows translations between default theories and causal theories and relates Pearl-style structural equation models to this logical perspective (Bochman 2023). Some recent works enrich causal models with additional structure that restricts admissible worlds and interventions. Beckers, Halpern, and Hitchcock study *causal models with constraints*, where constraints carve out impossible states that standard SCM semantics would otherwise allow (Beckers, Halpern, and Hitchcock 2023). This is closely aligned in spirit with our intervention–feasibility diagnosis, but orthogonal in methodology: constraints restrict which assignments count as admissible *globally*, whereas our R-SCM semantics restricts *agent-relative* counterfactual reasoning by requiring executable, resource-feasible traces. Halpern and Hitchcock enrich HP-style causation with a normality ordering over worlds, yielding graded judgments that prefer more normal contingencies (Halpern and Hitchcock 2015). This addresses a different axis of under-specification than the one targeted here. Our framework is driven by a different goal: blocking infeasible counterfactual reasoning by making admissibility constraints explicit and operational. Normality (or default) approaches *filter* candidate counterfactual contingencies based on typicality. In contrast, our goal to restrict the *domain* of relevant counterfactual antecedents to those witnessed by *enabled*, agent-relative manipulations. The induced enabled-intervention dynamics forms a planning-style reachability problem: actions are enabled primitive interventions and costs are resource-token consumption. The ideas discussed above combine naturally with our setting in that after restricting attention to feasible traces, a normality ordering can rank the remaining feasible alternatives.

Logical theories of responsibility in multi-agent systems typically take agency primitives (choices, strategies, game forms) as given and analyze responsibility via ability operators and their temporal or strategic refinements (Shi 2024; Parker, Grandi, and Lorini 2025). A recurring theme is that responsibility is sensitive to information and epistemic limitations, and this has been studied explicitly under imperfect information in strategic settings (Yazdanpanah et al. 2019). Our approach does not compete with these formalisms, rather, it provides an SCM-based substrate in which 'could have done otherwise' is witnessed operationally, with feasibility (and, when desired, information constraints) enforced at the level of enabled interventions. In algorithmic recourse, a closely related concern is that standard counterfactual-search formulations implicitly treat arbitrary feature settings as achievable, producing infeasible recommendations. This has motivated a shift from counterfactual explanations to actionable, intervention-aware recourse (Karimi, Scholkopf, and Valera 2021). R-SCMs can be viewed as a formalization of the same doctrine: recourse is evaluated only over counterfactual reasoning supported by feasible intervention traces 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 $\mathrm{do}(\cdot)$ 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 $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(X{=}x)]\,\varphi$ 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 an responsibility or recourse argument despite being infeasible for the agent. As mentioned previously, our core theoretical development (Section 4) re-

solves this by restricting agent-relative counterfactual reasoning to those witnessed by enabled, resource-consuming execution traces.

Reaction-time infeasibility. 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:

$$H \coloneqq A \wedge T \qquad B \coloneqq \neg H.$$

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 veracity). 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.

Causal recourse. In causal recourse, a denied subject is given changes that would flip a decision (Karimi, Scholkopf, and Valera 2021; Wang et al. 2025; von Kügelgen et al. 2022). Let $D \coloneqq \mathbb{I}[\mathrm{score}(X) \geq \tau]$ be a deterministic decision rule over features X. A standard objective searches for a small feature move ΔX such that $(\mathcal{M}, \vec{u}) \models [\mathrm{do}(X \coloneqq X + \Delta X)]\, (D{=}1)$. The failure mode is familiar: an *outcome-flipping* $\mathrm{do}(\cdot)$ antecedent is not a feasible option for the relevant agent, yet is treated as a legitimate. 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 (Simon 1955). 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 $V \coloneqq f_V(\mathrm{Pa}(V), \vec{U})$. 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 neighbors 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 $X \in \mathrm{dom}(I)$, each instance of X by the constant assignment $X \coloneqq I(X)$ in the corresponding structural equations. 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 $\mathcal{A} = \{1, \ldots, n\}$.

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 commutative binary operation (with associativity and unit laws holding whenever the relevant expressions are defined), and $e \in \mathcal{T}$ is the unit resource object. We write $t \otimes t' \downarrow$ to mean the composition is defined and $t \otimes t' \uparrow$ to mean it is undefined. Undefinedness represents a* resource clash*, and definedness represents compatibility. The induced availability preorder is $c \preceq t$ meaning there exists $r \in \mathcal{T}.\, (c \otimes r \downarrow \wedge\, c \otimes r = t)$. When $c \preceq t$, any such r is called a* residual *of paying c from t (residuals need not be unique). If $c \preceq t$, we 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* $\mathcal{C}_i \subseteq \mathcal{V}$, representing endogenous variables that are, in principle, directly actuated by i.

Remark 1. *When I is an intervention map with $\mathrm{dom}(I) \subseteq \mathcal{C}_i$, 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 (Winn 2012). 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 (Eberhardt 2009).

Concretely, each agent i is equipped with an observation map $\mathsf{obs}_i : \prod_{V \in \mathcal{V}} \mathcal{R}(V) \to \mathcal{O}_i$, and when the current cumulative intervention induce 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 in the same way — as additional gate constraints.

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 (which is enforced by executability), 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.* □

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 $X \in \mathrm{dom}(I)$, each instance of X by the constant assignment $X := I(X)$ in the corresponding structural equations. We write $\vec{v}_I := \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ for the induced endogenous valuation (which is well-defined for acyclic SCMs).* □

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 $x \in \mathcal{R}(X)$. Each action α^i has a resource-token requirement $\mathsf{cost}(\alpha) \in \mathcal{T}$ and a* gate predicate *$G_{\alpha^i}(\vec{u}, o_i, \rho(i), I) \in \{0, 1\}$, evaluated at context $\vec{u}$, observation o, resource store ρ, and current cumulative intervention map I. Intuitively, $G_{\alpha^i} = 1$ entails both resource availability and any additional feasibility constraints.*

We require that if $G_{\alpha^i}(\vec{u}, o_i, \rho(i), I) = 1$, then $\mathsf{cost}(\alpha) \preceq \rho(i)$. A primitive action $\alpha^i = (i, X{:=}x)$ is enabled *at s, written $s \models \mathsf{en}(\alpha)$, if $G_{\alpha^i}(\vec{u}, o_i, \rho(i), I) = 1$, and $X \notin \mathrm{dom}(I)$.* □

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

□

Definition 6 (Trace)**.** *A* trace *is a finite sequence of primitive actions $\sigma = \alpha_1 ; \cdots ; \alpha_m$. A trace is interpreted operationally via the one-step transition relation (Definition 5) and the induced cumulative intervention map (Definition 2).*

$\sigma = \alpha_1 ; \cdots ; \alpha_m$ is executable *from a state s if there exist states $s_0, \dots, s_m$ with $s_0 = s$ such that $s_{t-1} \xrightarrow{\alpha_t} s_t$ for each $t = 1, \dots, m$. We write $s \xrightarrow{\sigma} s_m$, and if $s_m = (\vec{u}, \rho_m, I_m)$ we write $I_\sigma := I_m$.* □

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

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

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, $\mathcal{C}_i \subseteq \mathcal{V}$ are controlled variables, Act *is a set of actions,* $\mathsf{cost} : \mathrm{Act} \to \mathcal{T}$ *assigns resource-token requirements, and each gate $G_\alpha(\vec{u}, o, \rho, I)$ determines admissibility of α at context $\vec{u}$, valuation $\vec{v}$, store $\rho : \mathcal{A} \to \mathcal{T}$ and cumulative intervention map I.* □

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 σ executable from s such that σ consists only of actions of agent i, and if $s \xrightarrow{\sigma} (\vec{u}, \rho', I_\sigma)$ then $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$* □

Following Halpern (Halpern 2016), we adopt the standard convention that an event E can be described by a propositional formula $varphi_E$. 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. Agent i is held *responsible* for event E with respect to an R-SCM ($\widehat{\mathcal{M}}$ if there exists a state s such that $(\widehat{\mathcal{M}}, s) \models \langle i \rangle \neg \varphi_E$. A recourse plan is simply an executable trace witnessing $\langle i \rangle \psi_D$, and minimal recourse plans correspond to resource-token minimal witnesses.

Definition 9 (Trivial feasibility)**.** *Given an R-SCM $\widehat{M}$, if for every primitive action $\alpha = (i, X{:=}x)$ with $X \in \mathcal{C}_i$ and $x \in \mathcal{R}(X)$, 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 $G_\alpha(\vec{u}, o, \rho, I) = 1$ and $\mathsf{cost}(\alpha) = e$ then feasibility is* trivial for the agent i. □

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 finite sequence $\sigma = \alpha_1 ; \cdots ; \alpha_m$ such that each action $\alpha_t = (i_t, X_t{:=}x_t)$ has $i_t \in C$ and $X_t \in \mathcal{C}_{i_t}$. A C-trace σ is* executable from *a state $s = (\vec{u}, \rho, I)$ iff there exist states $s_0, \dots, s_m$ with $s_0 = s$ such that $s_{t-1} \xrightarrow{\alpha_t} s_t$ for each $t = 1, \dots, m$ (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 σ, a C-trace executable from s such that $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. 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 and Feasible Recourse

Definition 11. *Fix an R-SCM $\widehat{\mathcal{M}}$, a state $s = (\vec{u}, \rho, I)$, an agent i, and an event B. A trace $\sigma = \alpha_1 ; \cdots ; \alpha_m$ is a* responsibility witness for preventing B by i at s *iff:*

1. *(*Potency*) $(\mathcal{M}, \vec{u}) \models B$.

2. *(*Feasible prevention*) If σ is executable from s with each $\alpha_t \in \sigma$ having the form $(i, X_t{:=}x_t)$ then $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \neg B$ where $s \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$*
3. *(*Token-minimality*) There is no executable i-trace σ' from s such that $(\mathcal{M}_{I_{\sigma'}}, \vec{u}) \models \neg B$ and $\mathsf{Cost}(\sigma') \prec \mathsf{Cost}(\sigma)$, where $\mathsf{Cost}(\sigma) := \bigotimes_{t=1}^{m} \mathsf{cost}(\alpha_t)$ and $\prec$ 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.* □

It is useful to make explicit how our feasibility witnesses correspond to the usual Halpern-Pearl (HP) clauses for actual causation (Halpern 2016). We emphasize that our goal is not to replace HP's causal semantics, but to *restrict the counterfactual alternatives* used for responsibility to those realizable by feasible procedures. HP's counterfactual dependence clause asks for an intervention (possibly under a contingency) under which the effect is prevented. Our refinement is to restrict the quantification domain of such interventions. Concretely, the *Prevention* clause in Definition 11 asks for a witness trace σ such that $s \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$ and $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \neg B$. HP imposes a minimality condition (typically subset-minimality over intervened variables). Here minimality condition is *resource-indexed*: we select witnesses that are resource-token cost minimal with respect to the resource-token preorder induced by $\otimes$. This is well-aligned with responsibility attribution, where the salient notion of 'smallest change' is typically budget rather than the cardinality of a variable set. When needed, one may add a secondary tie-breaker (e.g., trace length) among incomparable traces.

4.4 Worked Example: k-of-n Pause Switch

We use a k-of-n approval (multisig) pause switch (Itakura 1983; Micali, Ohta, and Reyzin 2001) as a running example because it characterizes a genuinely *coordination-dependent* alternative. It is a threshold control mechanism arising out of research in cryptography in which a system can be paused only after approvals from at least k out of n designated parties.

We show that no single signer can unilaterally pause a system to prevent some outcome (called loss), yet a sufficiently large coalition can — iff the relevant procedural and timing constraints keep the option live. We fix $n \geq 2$ and a threshold $k \in \{2, \ldots, n\}$. Let E denote that an exploit attempt is underway, let $A_1, \ldots, A_n$ 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. Consider the deterministic acyclic SCM with equations

$$\begin{array}{ll} Q := f_Q(A_1, \ldots, A_n) & f_Q(a_1, \ldots, a_n) = \mathbf{1}[\sum_{i=1}^{n} a_i \geq k] \\ P := f_P(E, Q) & f_P(e, q) = \mathbf{1}[\, e = 1 \wedge q = 1 \,] \\ B := f_B(E, P) & f_B(e, p) = \mathbf{1}[\, e = 1 \wedge p = 0 \,] \end{array}$$

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 situation satisfies $(\mathcal{M}, \vec{u}) \models E{=}1$ and $(\mathcal{M}, \vec{u}) \models A_i{=}0$ for all i, hence $Q{=}0$, $P{=}0$, and $B{=}1$.

For each signer $i \in \{1, \ldots, n\}$, 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 quantify 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 but they do not represent executable options: neither P nor Q are controlled by any signer.

To model the availability constraint, we introduce an endogenous variable $W \in \{0, 1\}$ indicating that the pause proposal is still live (this can be derived from other timing variables in the model). The approval action is enabled only while $W{=}1$ and the signer has sufficient resources: $G_{\alpha^i}(\vec{u}, o, \rho(i), I) = 1$ iff $(\vec{v}_I(W) = 1 \wedge (\mathsf{cost}(\alpha^i) \preceq \rho(i))$, where $o = \mathsf{obs}_i(\vec{v}_I)$ and $\vec{v}_I = \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$. Let the initial state be $s = (\vec{u}, \rho, \emptyset)$, and suppose the initial valuation $\vec{v}_0 := \mathrm{Sol}_{\mathcal{M}}(\vec{u})$ satisfies $\vec{v}_0(W) = 1$ (the window is open).

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 $k > 1$, $(\widehat{\mathcal{M}}, s) \not\models \langle i \rangle \neg B$. 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 and, along the relevant execution, the window remains open (i.e., each gate check sees $W = 1$). 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 $(\widehat{\mathcal{M}}, s) \models \langle C \rangle \neg B$. If instead $\vec{v}_0(W) = 0$ (the proposal window is closed), then no approval action is enabled and therefore $(\widehat{\mathcal{M}}, s) \not\models \langle C \rangle \neg B$ for every coalition C.

5 Meta-theory: 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 an 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 $s \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$, then I_σ 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 $|\sigma| \le |\mathcal{C}_i \setminus \mathrm{dom}(I)|$. More generally, if σ is an executable C-trace from s then $|\sigma| \le |\mathcal{C}_C \setminus \mathrm{dom}(I)|$ 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, $|\sigma| \le |\mathcal{C}_i \setminus \mathrm{dom}(I)|$. 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 $(\widehat{\mathcal{M}}, s) \models \langle C \rangle \varphi$ then $(\widehat{\mathcal{M}}, s) \models \langle C' \rangle \varphi$.* □

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 if $s \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$ then $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$. 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 action $\alpha = (i, X{:=}x)$ 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 $s \preceq_i s'$, $\rho(j) = \rho'(j)$ for all $j \neq i$ and $\rho(i) \preceq \rho'(i)$. Hence there exists some d with $\rho'_i = \rho_i \otimes d$. Let $c := \mathsf{cost}(\alpha)$. From $\rho(i) = c \otimes r$ and $\rho'(i) = \rho(i) \otimes d$ we obtain $\rho'(i) = (c \otimes r) \otimes d$. By associativity of $\otimes$ (where defined), $\rho'(i) = c \otimes (r \otimes d)$, so $c \preceq \rho'(i)$. Thus α's cost is payable from $\rho'(i)$ with residual $r' := r \otimes d$.

Moreover, $X \notin \mathrm{dom}(I)$ holds as the intervention map is the same in s and s', and by monotonicity of gates (Definition 12) we have, if $G_\alpha(\vec{u}, o, \rho, I) = 1$ then $G_\alpha(\vec{u}, o, \rho', I) = 1$. Therefore by Definition 5 there exists a $s'_1 = (\vec{u}, \rho'_1, I_1)$ with the *same* intervention-map update $I_1 = I \cup \{X \mapsto x\}$, and with $\rho'_1(i) = r'$ and $\rho'_1(j) = \rho'(j)$ for $j \neq i$.

By construction, s_1 and s'_1 have the same $(\vec{u}, I_1)$, and $\rho_1(j) = \rho'_1(j)$ for all $j \neq i$ while $\rho_1(i) = r \preceq r' = \rho'_1(i)$. Hence $s_1 \preceq_i s'_1$.

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. The formula $\langle i \rangle \varphi$ holds ex-

actly 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 $\mathrm{dom}(J) \subseteq \mathcal{C}_i$ 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 $\alpha_t = (i, X_t{:=}x_t)$ with $X_t \in \mathcal{C}_i$), and if $s \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$ then $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$.

Let $J := I_\sigma$. Therefore, every assignment recorded in J targets a variable in $\mathcal{C}_i$, hence $\mathrm{dom}(J) \subseteq \mathcal{C}_i$, and so, J is an i-control intervention (cf. Remark 1). 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 $\mathrm{dom}(J) \subseteq \mathcal{C}_i$ such that $(\mathcal{M}_J, \vec{u}) \models \varphi$. Let us enumerate $\mathrm{dom}(J) = \{X_1, \ldots, X_m\}$ (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 $t = 1, \ldots, m$ 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 $t \in \{1, \ldots, m\}$ 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 $\alpha_t = (i, X_t{:=}J(X_t))$. To apply the one-step rule (Definition 5) it suffices to check $X_t \notin \mathrm{dom}(I_{t-1})$, which holds by construction; and whether $G_{\alpha_t}(\vec{u}, o_{t-1}, \rho_{t-1}, I_{t-1}) = 1$

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 $G_{\alpha_t}(\vec{u}, o_{t-1}, \rho_{t-1}, I_{t-1}) = 1$. Trivial feasibility further provides $\mathsf{cost}(\alpha_t) = e$. 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 $\{X_1, \ldots, X_m\} = \mathrm{dom}(J)$, 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* FEASIBILITY$_k$ *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 if $s_0 \xrightarrow{\sigma} (\vec{u}, \rho_\sigma, I_\sigma)$ then $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$.* □

We specify k in unary in 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 (Halpern 2016) and (Eiter and Lukasiewicz 2002; Aleksandrowicz et al. 2017) for a discussion of computational complexity results for the Halpern-Pearl framework. We follow (Eiter and Lukasiewicz 2002) 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. Further assume that each structural function f_V and each gate predicate G_α is computable in time polynomial in the input size, and that resource operations are polynomial-time decidable: given c, t, one can decide whether $c \preceq t$ and, when so, compute a residual r with $c \otimes r = t$. With k given in unary,* FEASIBILITY$_k$ *is NP-complete.* □

Proof. We first show that FEASIBILITY$_k \in$ NP. There exists a nondeterministic algorithm which guesses a trace $\sigma = \alpha_1; \cdots; \alpha_m$ with $m \leq k$. $|\sigma|$ is polynomial in the input size when k is given in unary, and under set-once discipline we may also assume $m \leq |\mathcal{C}_C|$). It then verifies executability step-by-step from $s_0 = (\vec{u}, \rho, \emptyset)$. For each prefix I_t induced by $\alpha_1; \cdots; \alpha_t$, it computes the unique valuation $\vec{v}_t = \mathrm{Sol}_{\mathcal{M}_{I_t}}(\vec{u})$ by evaluating the acyclic structural equations in some topological order (which is polynomial time since each f_V is poly-time evaluable), checks the gate predicate $G_{\alpha^i_{t+1}}(\vec{u}, \mathsf{obs}_{i_{t+1}}(\vec{v}_t), \rho_t(i), I_t) = 1$ (which is polynomial-time by assumption), and updates the resource store by computing a residual after paying $\mathsf{cost}(\alpha_{t+1})$ (poly-time by the resource assumptions). After m steps it checks $(\mathcal{M}_{I_m}, \vec{u}) \models \varphi$. Hence FEASIBILITY$_k \in$ NP.

To show NP-hardness, we establish a poly-time reduction from FEASIBILITY$_k$ to SAT. Let ψ be a CNF formula over propositional variables $x_1, \ldots, x_n$. We construct in polynomial time an instance of FEASIBILITY$_k$ that is a YES-instance iff ψ is satisfiable. We begin by constructing an SCM with Boolean endogenous variables $X_1, \ldots, X_n, Y$, where $X_j = 1$ (respectively 0) encodes the truth value of x_j (respectively $\neg x_j$). Let $Y := \psi(X_1, \ldots, X_n)$, viewing ψ as the Boolean function obtained by interpreting each x_j as the value of X_j. We take $\mathcal{U} = \emptyset$ and define an acyclic SCM $\mathcal{M}$ by setting, for each $j \in \{1, \ldots, n\}$, $X_j := 0$ and $Y := \psi(X_1, \ldots, X_n)$, where ψ is evaluated as a Boolean circuit on the valuation of $(X_1, \ldots, X_n)$.

Now let us consider a single agent i with $\mathcal{C}_i = \{X_1, \ldots, X_n\}$, and for each j and $b \in \{0, 1\}$, the primitive action $(i, X_j{:=}b)$. Let all actions be always enabled and let all costs be neutral ($\mathsf{cost}(\alpha) = e$), with any initial store ρ. Let the initial state be $s_0 = (\rho, \emptyset)$, and set the goal event to $\varphi \equiv (Y = 1)$, and the horizon to $k := n$.

If ψ is satisfiable, let $a_1, \ldots, a_n \in \{0, 1\}$ be a satisfying assignment. Consider the trace $\sigma := (i, X_1{:=}a_1)\ ;\ \cdots\ ;\ (i, X_n{:=}a_n)$. Every step is executable, and by construction

$I_\sigma(X_j) = a_j$ for all j. Hence in the intervened model $\mathcal{M}_{I_\sigma}$ we have $Y = \psi(a_1, \ldots, a_n) = 1$, so $(\mathcal{M}_{I_\sigma}) \models \varphi$. Therefore this instance is a YES-instance.

Conversely, suppose the instance is a YES-instance. Then there exists an executable trace σ of length at most n such that $(\mathcal{M}_{I_\sigma}) \models (Y = 1)$. Let us define an assignment $a_1, \ldots, a_n \in \{0, 1\}$ by $a_j := I_\sigma(X_j)$ if $X_j \in \mathrm{dom}(I_\sigma)$ and $a_j := 0$ otherwise. Under the set-once discipline, each X_j is assigned at most once along σ, so this is well-defined. In $\mathcal{M}_{I_\sigma}$ the variables X_j evaluate to a_j, and since $Y := \psi(X_1, \ldots, X_n)$ we obtain $\psi(a_1, \ldots, a_n) = 1$. Hence ψ is satisfiable. This reduction is polynomial-time, and therefore FEASIBILITY$_k$ is NP-complete. □

6 Conclusion

We have characterized an *intervention–feasibility gap* in the Halpern-Pearl (HP) structural-model of actual causation. The proposed R-SCMs (Definition 7) address this by a separation of concern. On the one hand, *dynamics* — what would follow from a hypothetical intervention — is still determined entirely by the HP structural equations in the intervened model $\mathcal{M}_I$. On the other hand, *procedural feasibility* — what an agent or coalition can actually bring about under operative time, authority, coordination, and policy constraints — is made explicit via agent-specific controlled variables together with operational enablement conditions (gates) and consumable resources.

An R-SCM can be viewed as a pair of orthogonal layers that interact only through the evaluation points induced by cumulative interventions. Given a context $\vec{u}$ and an intervention map I, the *dynamics layer* is exactly standard Halpern-Pearl evaluation in the intervened SCM, yielding the unique valuation $\vec{v}_I = \mathrm{Sol}_{\mathcal{M}_I}(\vec{u})$ and hence truth of events $(\mathcal{M}_I, \vec{u}) \models \varphi$. Independently, the *feasibility layer* is the labeled transition system on states $s = (\vec{u}, \rho, I)$ generated by the one-step rule (Definition 5), where enabledness of a primitive action $\alpha = (i, X{:=}x)$ is decided by the gate predicate evaluated at the agent's information $o = \mathsf{obs}_i(\vec{v}_I)$ together with the current store ρ. This makes responsibility (and thus preventability) and recourse claims audit-capable — one exhibits a concrete enabling sequence — while remaining compatible with traditions that begin from an action signature with preconditions and effects (Tran and Baral 2004; Hopkins and Pearl 2007; Batusov and Soutchanski 2018; LeBlanc, Balduccini, and Vennekens 2019; Liu and Belle 2025).

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 $\otimes$ — the same mathematical device used to model disjointness in BI and separation-logic semantics (Ishtiaq and O'Hearn 2001; Reynolds 2002; Pym 2019; Gheorghiou and Pym 2023). In consequence, feasibility witnesses are *semantic certificates*: $(\widehat{\mathcal{M}}, s) \models \langle C \rangle \varphi$ is witnessed by an explicit executable trace σ such that $(\mathcal{M}_{I_\sigma}, \vec{u}) \models \varphi$.

Because the witness is already operational, it is natural to internalize it in a proof system. We hypothesize that a BI-inspired system (Ishtiaq and O'Hearn 2001; Pym 2019; Gheorghiou and Pym 2023) can internalize this witness as a two-context judgment with a structural fact context and a substructural resource store, and derive a *frame* principle stating that resources not touched by a certificate can be carried along unchanged (when compatible): any derivation yields a concrete trace σ with $(\vec{u}, \rho, I) \xrightarrow{\sigma} (\vec{u}, \rho', I')$ and $(\mathcal{M}_{I'}, \vec{u}) \models \varphi$. Such a calculus is conjectured to support a BI-style frame principle — resources not mentioned by a certificate are irrelevant and may be carried along unchanged — thereby aligning with interface-level systems modeling where procedures and controls are explicit design elements (Chakraborty, Caulfield, and Pym 2025b; Chakraborty, Caulfield, and Pym 2025a). Such certificates suggest a lightweight tooling path for feasibility and preventability claims, to be incorporated with machine-checkable proofs, and verified independently (or synthesized via SAT/SMT back-ends) as artifacts that can be audited.

We focus on deterministic acyclic SCMs and bounded traces to keep counterfactual evaluation unambiguous and obtain crisp complexity bounds. Natural extensions include richer (possibly ownership-tagged) resource algebras for shared or global exclusives, non-commutative composition for ordered workflows, and epistemic or probabilistic constraints so feasibility is evaluated relative to what agents can observe or infer. The conceptual recipe remains the same: an SCM $\mathcal{M}$ constitutes the causal substrate while additional structure refines the admissibility layer that governs which counterfactual alternatives count as feasible.

Acknowledgements

Blinded.

References

Alechina, N.; Logan, B.; Nga, N. H.; and Rakib, A. 2010. Resource-bounded Alternating-time Temporal Logic. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1*, AAMAS '10, 481–488. International Foundation for Autonomous Agents and Multiagent Systems.

Aleksandrowicz, G.; Chockler, H.; Halpern, J. Y.; and Ivrii, A. 2017. The Computational Complexity of Structure-based Causality. *J. Artif. Int. Res.* 58(1):431–451.

Baral, C., and Gelfond, M. 2000. *Reasoning Agents in Dynamic Domains*. Boston, MA: Springer US. 257–279.

Batusov, V., and Soutchanski, M. 2018. 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.

Beckers, S.; Halpern, J. Y.; and Hitchcock, C. 2023. Causal Models with Constraints. In *Proceedings of the Second Conference on Causal Learning and Reasoning*, volume 213 of *Proceedings of Machine Learning Research*, 866–879. PMLR.

Bochman, A. 2023. Default Logic as a Species of Causal Reasoning. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*.

Chakraborty, P.; Caulfield, T.; and Pym, D. 2025a. A Logic for Resource-sensitive Coalition Games. In *16th International Conference, GameSec 2025*. LNCS 16223:61–80, Springer.

Chakraborty, P.; Caulfield, T.; and Pym, D. 2025b. Local Causal Reasoning in Multiagent Systems (Extended Abstract). In *Proc. The 2nd International Workshop on Causality, Agents and Large Models (CALM), in press*, Communications in Computer and Information Science series. Springer.

Chockler, H., and Halpern, J. Y. 2003. Responsibility and Blame: A Structural-model Approach. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, 147–153. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

de Sio, F. S., and van den Hoven, J. 2018. Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI* 5.

Eberhardt, F. 2009. Introduction to the Epistemology of Causation. *Philosophy Compass* 4(6):913–925.

Eiter, T., and Lukasiewicz, T. 2002. Complexity results for structure-based causality. *Artificial Intelligence* 142(1):53–89.

Feldman, M. S., and Pentland, B. T. 2003. Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Administrative Science Quarterly* 48(1):94–118.

Finzi, A., and Lukasiewicz, T. 2002. Structure-based Causes and Explanations in the Independent Choice Logic. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, UAI'03, 225–323.

Galles, D., and Pearl, J. 1997. Axioms of Causal Relevance. *Artificial Intelligence* 97(1):9–43.

Gheorghiou, A. V., and Pym, D. J. 2023. Semantical Analysis of the Logic of Bunched Implications. *Studia Logica* 111:525–571.

Halpern, J. Y., and Hitchcock, C. 2015. Graded Causation and Defaults. *British Journal for the Philosophy of Science* 66(2):413–457.

Halpern, J. Y., and Kleiman-Weiner, M. 2018. 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.

Halpern, J. Y. 2016. *Actual Causality*. Cambridge, MA: The MIT Press.

Hopkins, M., and Pearl, J. 2007. Causality and Counterfactuals in the Situation Calculus. *Journal of Logic and Computation* 17(5):939–953.

Ishtiaq, S. S., and O'Hearn, P. W. 2001. BI as an Assertion Language for Mutable Data Structures. *ACM SIGPLAN Notices* 36(3):14–26.

Itakura, K. 1983. A Public-key Cryptosystem Suitable for Digital Multisignature. *NEC research and development* 71:1–8.

Karimi, A.-H.; Scholkopf, B.; and Valera, I. 2021. Algorithmic Recourse: from Counterfactual Explanations to Interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 353–362. New York, NY, USA: Association for Computing Machinery.

LeBlanc, E.; Balduccini, M.; and Vennekens, J. 2019. Explaining Actual Causation via Reasoning About Actions and Change. In *Logics in Artificial Intelligence*, 231–246. Springer International Publishing.

Liu, D., and Belle, V. 2025. What Is a Counterfactual Cause in Action Theories? In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, 2627–2629. International Foundation for Autonomous Agents and Multiagent Systems.

Lorini, E.; Longin, D.; and Mayor, E. 2013. A Logical Analysis of Responsibility Attribution: Emotions, Individuals and Collectives. *Journal of Logic and Computation* 24(6):1313–1339.

Micali, S.; Ohta, K.; and Reyzin, L. 2001. Accountable-subgroup Multisignatures: Extended Abstract. In *Proceedings of the 8th ACM Conference on Computer and Communications Security*, 245–254. New York, NY, USA: Association for Computing Machinery.

Nguyen, H. N.; Alechina, N.; Logan, B.; and Rakib, A. 2019. Probabilistic Resource-bounded Alternating-time Temporal Logic. *Artificial Intelligence* 275:182–222.

Parker, T.; Grandi, U.; and Lorini, E. 2025. Responsibility in a Multi-value Strategic Setting. In *Multi-Agent Systems: 21st European Conference, EUMAS 2024, Dublin, Ireland, August 26–28, 2024, Proceedings*, 138–158. Cham: Springer Nature Switzerland.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. USA: Cambridge University Press, 2nd edition.

Poole, D. 1997. The Independent Choice Logic for Modelling Multiple Agents under Uncertainty. *Artificial Intelligence* 94(1):7–56. Economic Principles of Multi-Agent Systems.

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

Pym, D. 2019. Resource semantics: logic as a modelling technology. *ACM SIGLOG News* 6(2):5–41.

Reiter, R. 2001. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. The MIT Press.

Reynolds, J. C. 2002. Separation Logic: A Logic for Shared Mutable Data Structures. In *Proceedings of the 17th IEEE Symposium on Logic in Computer Science (LICS)*, 55–74. IEEE.

Shi, Q. 2024. Responsibility in Extensive Form Games. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press.

Simon, H. A. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* 69(1):99–118.

Simon, H. A. 2019. *The sciences of the artificial (3rd ed.)*. The MIT Press.

Tran, N., and Baral, C. 2004. Encoding Probabilistic Causal Model in Probabilistic Action Language. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, 305–310. AAAI Press.

Ustun, B.; Spangher, A.; and Liu, Y. 2019. Actionable Recourse in Linear Classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 10–19. Association for Computing Machinery.

von Kügelgen, J.; Karimi, A.-H.; Bhatt, U.; Valera, I.; Weller, A.; and Schölkopf, B. 2022. On the Fairness of Causal Algorithmic Recourse. *Proceedings of the AAAI Conference on Artificial Intelligence* 36(9):9584–9594.

Wang, H.; Zou, H.; Zhou, X.; Wang, S.; Yang, W.; and Cui, P. 2025. Learning Feasible Causal Algorithmic Recourse: A Prior Structural Knowledge Free Approach. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, 4507–4518. New York, NY, USA: Association for Computing Machinery.

Winn, J. 2012. Causality with Gates. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, 1314–1322. La Palma, Canary Islands: PMLR.

Yazdanpanah, V.; Dastani, M.; Jamroga, W.; Alechina, N.; and Logan, B. 2019. Strategic Responsibility Under Imperfect Information. In *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems*, 592–600. International Foundation for Autonomous Agents and Multiagent Systems.